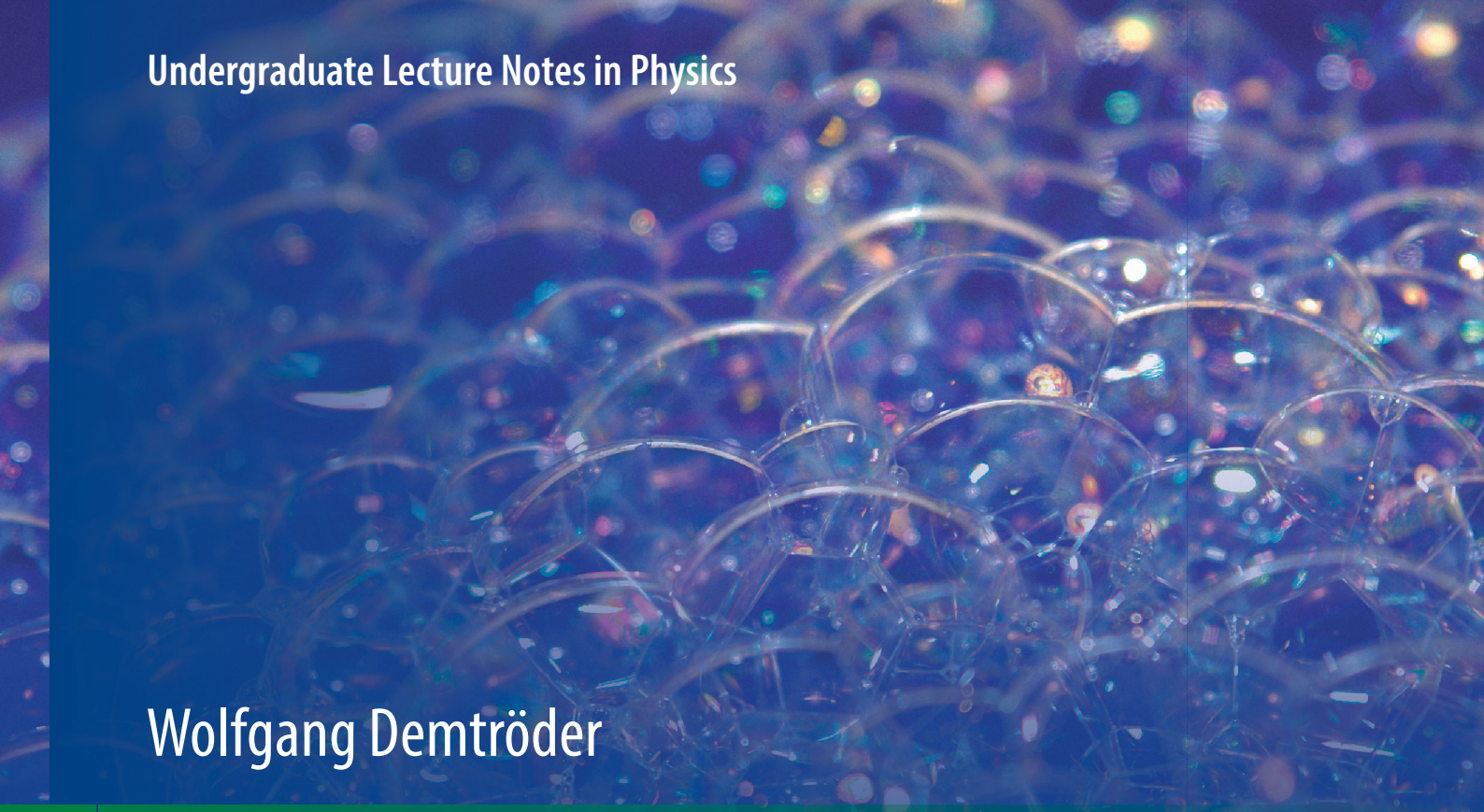
Undergraduate Lecture Notes in Physics



Wolfgang Demtröder

Astrophysics

Springer

Undergraduate Lecture Notes in Physics

Series Editors

Neil Ashby, University of Colorado, Boulder, CO, USA

William Brantley, Department of Physics, Furman University, Greenville, SC, USA

Matthew Deady, Physics Program, Bard College, Annandale-on-Hudson, NY, USA

Michael Fowler, Department of Physics, University of Virginia, Charlottesville, VA, USA

Morten Hjorth-Jensen, Department of Physics, University of Oslo, Oslo, Norway

Michael Inglis, Department of Physical Sciences, SUNY Suffolk County Community College, Selden, NY, USA

Barry Luokkala, Department of Physics, Carnegie Mellon University, Pittsburgh, PA, USA

Undergraduate Lecture Notes in Physics (ULNP) publishes authoritative texts covering topics throughout pure and applied physics. Each title in the series is suitable as a basis for undergraduate instruction, typically containing practice problems, worked examples, chapter summaries, and suggestions for further reading.

ULNP titles must provide at least one of the following:

- An exceptionally clear and concise treatment of a standard undergraduate subject.
- A solid undergraduate-level introduction to a graduate, advanced, or non-standard subject.
- A novel perspective or an unusual approach to teaching a subject.

ULNP especially encourages new, original, and idiosyncratic approaches to physics teaching at the undergraduate level.

The purpose of ULNP is to provide intriguing, absorbing books that will continue to be the reader's preferred reference throughout their academic career.

Wolfgang Demtröder

Astrophysics

Springer

Wolfgang Demtröder
Demtröder, Germany

ISSN 2192-4791 ISSN 2192-4805 (electronic)
Undergraduate Lecture Notes in Physics
ISBN 978-3-031-22133-0 ISBN 978-3-031-22135-4 (eBook)
https://doi.org/10.1007/978-3-031-22135-4

Translation from the German language edition: “Experimentalphysik 4” by Wolfgang Demtröder, © Springer-Verlag GmbH Deutschland 2017. Published by Springer Spektrum. All Rights Reserved.

© Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This textbook on Astronomy and Astrophysics is an enlarged and updated version of the German textbook, I wrote some years ago. It is addressed to science students, in order to awaken their interest in the fascinating field of Astronomy. The author believes that all people, educated in science, should know at least the basic facts of Astronomy, since they not only can give an explanation about the common events, we are daily confronted with, such as day and night, summer and winter. or sun and moon eclipses, but they also inform about such questions as "What are stars, how long do they live and how many exist in the universe?" There are more fundamental questions, such as "Where do we come from", is the existence of life originated on Earth or was it brought to Earth by meteorites? Are we the only intelligent beings in the universe or are there other and may be even more advanced civilisations on other planets? If they exist, how large is the chance to communicate with them?

Although up to now there is no definitive answer to these latter questions, it is interesting to learn, how can we try to answer them and what are the experimental possibilities to observe extrasolar habitable planets.

This textbook tries to give the basic information about the way astronomers find out about the different stages of the development of stars. Besides coordinate systems, necessary to define the position of stars, the main insruments in Astronomy, such as modern telescopes, used in the different spectral ranges, detectors and space probes are discussed. The interesting question, where the energy comes from, that stars emit during their lifetimes, and information about the different final stages of stars are outlined in some more detail.

The author does hope to find many interesting readers and he would appreciate their comments and suggestions for improvement of the presentation.

Most of the references, collected at the end of each chapter are besides general books on astronomy and special articles on the internet which are faster accessible to the reader than articles in journals.

Many thanks go to colleagues who gave their permission for pictures. Several photographs are taken from the different NASA missions and I thank for the general allowance to reproduce them. I also thank Mrs. Ute Heuser, Springer Verlag, for organizing the editing of the book and my wife Harriet for her patient support during the time it took to write this book.

Kaiserslautern, Germany
July 2024

Wolfgang Demtröder

Contents

1 **Introduction** 1
1.1 History of Astronomy and Astrophysics 2
1.2 Survey on the Concept of This Textbook 4
References 5

2 Basic Principles and Techniques of Experimental Astronomy 7
2.1 Astronomical Coordinate Systems 7
2.1.1 The Horizontal System 7
2.1.2 The Equatorial Coordinate System 8
2.1.3 The Ecliptical System 10
2.1.4 The Galactic Coordinate System 10
2.2 Temporal Changes of the Coordinates 11
2.3 Definitions of Astronomical Times 12
2.4 The UTC is Nowadays Worldwide Used 12
2.5 Observation of Stars 12
2.6 Telescopes 14
2.6.1 Light Gathering Power of Telescopes 14
2.6.2 Angular Magnification of Telescopes 16
2.6.3 Different Types of Telescopes 16
2.6.4 Angular Resolution of Telescopes 19
2.6.5 Active Optics 19
2.6.6 Adaptive Optics 20
2.6.7 Large Telescopes 20
2.6.8 Telescope Mounts 22
2.6.9 Radio Telescopes 23
2.6.10 Optical Interferometry in Astronomy 25
2.6.11 X-ray Telescopes 28
2.6.12 Detectors for Cosmic Gamma Radiation 29
2.6.13 High Energy Cosmic Radiation 31
2.7 Detectors for Gravitational Waves 32
Summary 33
Problems 34
References 35

3 Measuring Techniques and Detectors 37
3.1 Measuring Techniques in Astronomy 37
3.1.1 Parallax 37
3.1.2 Aberration 38
3.1.3 Refraction in the Atmosphere 39
3.2 Measurements of Star Distances 39
3.2.1 Geometrical Methods 39
3.2.2 Time of Flight Measurements with Laser Pulses 40

3.2.3 Trigonometric Parallax Measurements . . . 42
3.2.4 Astrometric Satellite Gaia . . . 42
3.2.5 Other Methods for Measurements of Star Distances . . . 44
3.2.6 Supernovae as Distance Candles . . . 45
3.3 Apparent and Absolute Magnitudes of Stars . . . 45
3.4 Measurements of the Spectral Distribution of Star Radiation . . . 47
Summary . . . 48
Problems . . . 48
References . . . 49

4 Our Solar System . . . 51
4.1 Introduction . . . 51
4.2 Kepler's Laws of Planetary Motion . . . 52
4.2.1 Elliptical Planetary Orbits and Kepler's First Law . . . 52
4.2.2 Second Kepler-Law . . . 54
4.2.3 Orbital Elements of the Planets . . . 56
4.2.4 The Periods of the Planets . . . 58
4.2.5 Size, Mass and Mean Density of the Planets . . . 59
4.2.6 The Names of the Planets . . . 60
4.2.7 Energy Budget of the Planets . . . 61
4.3 The Inner Planets and Their Moons . . . 63
4.3.1 Mercury . . . 63
4.3.2 Venus . . . 64
4.3.3 The Earth . . . 65
4.3.4 Mars . . . 72
4.4 The Outer Planets . . . 74
4.4.1 Jupiter and Its Moons . . . 75
4.4.2 Saturn . . . 78
4.4.3 The Outermost Planets . . . 81
4.5 Small Celestial Bodies in the Solar System . . . 82
4.5.1 Dwarf Planets . . . 82
4.5.2 Planetoids . . . 85
4.5.3 Comets . . . 88
4.5.4 Meteors and Meteorites . . . 90
4.6 The Sun as Stationary Star . . . 92
4.6.1 Mass, Size, Density and Luminosity of Our Sun . . . 92
4.6.2 Mean Values of Temperature and Pressure Inside the Sun . . . 94
4.6.3 Radial Dependence of Pressure, Density and Temperature . . . 95
4.6.4 Energy Production Inside the Sun . . . 97
4.6.5 The Problem with the Sun-Neutrinos . . . 100
4.6.6 The Energy Transport in the Sun . . . 102
4.6.7 The Photosphere . . . 104
4.6.8 Chromosphere and Corona . . . 108
4.7 The Active Sun . . . 111
4.7.1 Sunspots . . . 111
4.7.2 The Magnetic Field of the Sun . . . 113
4.7.3 Facula, Flares and Protuberances . . . 115
4.7.4 The Pulsating Sun; Helio-Seismology . . . 116
Summary . . . 117
Problems . . . 119
References . . . 119

5 Birth, Lifetime and Death of Stars . . . 121
5.1 The Stars in Our Neighbourhood . . . 122
5.2 Direct Measurements of Star Radii . . . 124
5.3 Binaries and the Determination of Masses and Radii of Stars . . . 125
5.3.1 Visual Binaries . . . 126
5.3.2 Spectroscopic Binaries . . . 127
5.3.3 Photometric Binaries . . . 127
5.3.4 Spectral Types of Stars . . . 128
5.3.5 Hertzsprung-Russel-Diagram . . . 128
5.4 The Birth of Stars . . . 130
5.4.1 The Jeans Criterion . . . 130
5.4.2 The Formation of Proto-Stars . . . 134
5.4.3 The Influence of Rotation on Contracting Gas Clouds . . . 135
5.4.4 The Progression of a Star into the Main Sequence of the Hertzsprung-Russel Diagram . . . 136
5.5 The Stable Stage on the Main Sequence . . . 136
5.5.1 The Dependence of Luminosity and Lifetime of Stars on their Mass . . . 137
5.5.2 Energy Production in Stars on the Main Sequence . . . 138
5.6 Star Development After Leaving the Main Sequence . . . 140
5.6.1 Development of Stars with small Masses . . . 141
5.6.2 Development of Stars with Medium Masses . . . 141
5.6.3 The Development of Stars with large Masses and the Synthesis of Elements Up to Iron . . . 142
5.7 End-Stages of Stars . . . 145
5.7.1 Degenerate Star-Material . . . 145
5.7.2 Equation of State for Degenerate Matter . . . 146
5.7.3 White Dwarfs . . . 147
5.7.4 Neutron Stars . . . 150
5.7.5 Black Holes . . . 155
5.8 Quasars . . . 161
5.9 Observable Phenomena During the Final Stage of Stars . . . 161
5.9.1 Pulsating Variable Stars . . . 162
5.9.2 When Stars Steal Mass . . . 165
5.9.3 Novae . . . 166
5.9.4 Supernovae . . . 166
5.9.5 Planetary Nebula and Supernova Remnants . . . 169
5.10 Summarizing the Star Development . . . 170
5.11 Thought Provoking . . . 172
Summary . . . 173
Problems . . . 174
References . . . 174

6 The Development and Present Structure of the Universe . . . 177
6.1 Experimental Proofs of a Finite Expanding Universe . . . 178
6.1.1 The Expansion of the Universe . . . 178
6.1.2 Cosmic Background Radiation . . . 178
6.1.3 Abundance Ratios of the Light Elements . . . 180
6.1.4 The Olbert's Paradoxon . . . 180
6.1.5 Homogenity of the Universe . . . 181
6.2 The Metric of a Curved Space . . . 182
6.3 The Standard Model . . . 184

6.3.1 Radiation Dominated and Matter Dominated Universe 184
6.3.2 Hubble Parameter and Critical Density 185
6.3.3 The Early Phase of the Universe 188
6.3.4 The Synthesis of the Light Elements 191
6.3.5 Transition from the Radiation-Dominated to the Mass-Dominated Universe 192
6.3.6 Decoupling of Matter and Radiation 192
6.3.7 The Formation of Globular Clusters and Galaxies 193
6.3.8 The Age of the Universe 193
6.3.9 Friedmann Equations 194
6.4 Formation and Structure of Galaxies 201
6.4.1 Dark Matter and Formation of Spatial Structure of the Matter Density 202
6.4.2 What Are Galaxies? 203
6.4.3 Different Types of Galaxies 203
6.4.4 Active Galaxies 206
6.4.5 Galaxy Clusters and Super Cluster 208
6.4.6 Colliding Galaxies 210
6.4.7 Gravitational Lenses 210
Summary 211
Problems 212
References 212

7 Our Home Galaxy: The Milky Way 215
7.1 Introduction 215
7.2 Stellar Statistics and Star Populations 215
7.3 Motions of the Stars in Our Neighbourhood 217
7.4 The Differential Rotation of the Galactic Disc 218
7.5 The Spiral Arms 221
7.6 Globular Clusters 222
7.7 Open Star Clusters 224
7.8 The Center of Our Galaxy 225
7.9 The Black Hole in the Center of Our Galaxy 225
7.10 The Dynamics of Our Galaxy 227
7.11 The Space Between the Stars; Interstellar Matter 227
Summary 231
Problems 231
References 231

8 The Dark Universe 233
8.1 Dark Matter 233
8.2 Dark Energy 235
Summary 236
Problems 236
References 237

9 The Evolution of Our Solar System 239
9.1 The Origin of the Elements 239
9.2 The Origin and Development of Our Solar System 241
9.2.1 Collaps of the Rotating Gas Cloud 241
9.2.2 Formation of Planetesimals 243
9.3 Separation of Gases and Solid Matter 244
9.4 The Age of Our Solar System 245
9.5 Other Solar Systems and Exoplanets 248

9.5.1 Observation Methods . . . 248
9.5.2 Conditions for the Development of Intelligent Life . . . 249
9.6 The Development of the Earth . . . 250
9.6.1 The Separation of Core and Earth Mantle . . . 250
9.6.2 The Earth's Crust . . . 252
9.6.3 Volcanic Activity . . . 253
9.6.4 Formation of the Oceans . . . 253
9.6.5 The Origin of the Earth Atmosphere . . . 254
9.6.6 The Development of Life on Earth . . . 256
Summary . . . 257
Problems . . . 257
References . . . 258

Solutions of Problems . . . 259

Time Table of Astronomy . . . 277

Index . . . 281

Introduction

1

Contents

1.1 History of Astronomy and Astrophysics 2
1.2 Survey on the Concept of This Textbook. 4
References. 5

Astronomy is a very old science. The observation of stars and planets and the first determination of their locations in the sky, their motions during the year and possible changes in time date back to the Babylonians, several thousand years before Christ. The Babylonians [1] as well as the Chinese and the Egyptians made exact observations of stars and planets. Their scientific work, however, was restricted to the determination of positions and motions of the celestial bodies but they could not find out, that stars are huge hot gas balls, and they did not know about the true distance of stars and planets from the earth, and why they move (Fig. 1.1).

The thorough understanding of stars started only in the nineteenth century when astronomy was extended to astrophysics, where the results of laboratory investigations, the development of atomic and nuclear physics and the invention of large telescopes allowed more detailed observations and their interpretations.

Nowadays stars are no longer regarded as eternal constant objects, which do not change, but it is recognized that stars are created, undergo a development and finally die. Many astronomers, Physicists and Chemists have contributed to the progress of our understanding of the development of stars and all other celestial objects and also of the history of our universe.

Mankind has always tried to enlarge the confines of perception in order to learn what lies behind the horizon of our present knowledge. Nuclear-and Particle- Physics provide an impressive example, how these borders of our knowledge have been pushed into the subatomic range with dimensions smaller than that of atomic nuclei.

The relevance of astronomy extends beyond the physical concept into a frontier region between natural science and Philosophy. This becomes even more obvious when we regard the importance of High Energy Physics for the understanding of the development of our universe. The *Big-Bang Model* and conceptions of an *inflationary universe* and the formation of the chemical elements in stars are essentially based on information from nuclear- and particle physics.

Astrophysics is based on insights into atomic physics, nuclear and high energy physics and Astronomy. It postulates instead of a static unchanging eternal universe a dynamic model that changes in time, where stars are born, evolve as long as their energy source is active and die finally in gigantic explosions. Galaxies as the accumulation of millions to billions of stars are formed, develop and may merge possibly with other galaxies. Astrophysics tries to understand the development of our universe, the formation of stars and their internal structure, the origin of the chemical elements we find today on Earth and the possible future of the universe [1–3].

Recent measurements of super-novae in far distant galaxies seem to indicate that our universe is expanding with positive acceleration, contrary to the opinion before, where the expansion should be retarded because of the attractive gravitational force on expanding galaxies by the mass of the universe. This accelerated expansion demands a repulsive force, attributed to the "**dark energy**", which nobody can explain up to now.

Questions regarding the beginning of our universe, the time of the Big Bang and its possible future effects, problems of Philosophy and Religion, which also concern the self-concept of human beings, their position in the world, questions like "Are we unique in the universe, or are there other intelligent beings on other planets?" In order to answer such questions intense research on extra-solar planets has started recently. Meanwhile, more than 4000 extra-solar planets have been discovered, some of them show similar conditions as on Earth [2–8]

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_1

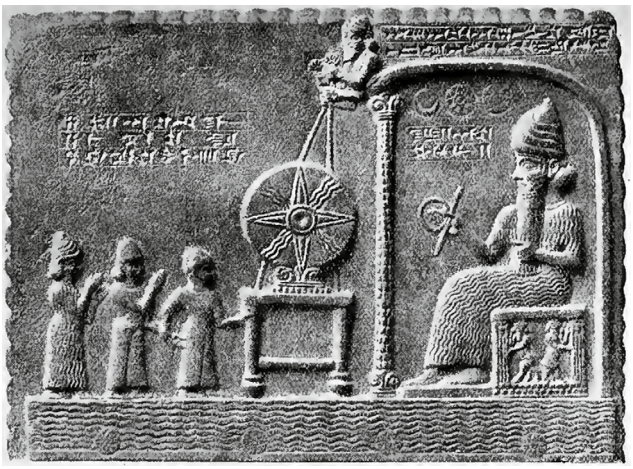


Fig. 1.1 The Mesopotamian King Hammurabi with astronomical symbols of the sun (https://www.bing.com/images/search)

Besides these epistemic questions, there are many aspects of practical applications of the results of Astrophysics to urgent problems on Earth. One example is the detailed study of the atmospheres of Venus and Mars, which has given much information about our own atmosphere, its origin, its development and its stability against manmade perturbations. At which CO_2 concentrations does it become unstable and what is the temperature rise then? The deterrent example of the Venus atmosphere (pressure p = 90 atmospheres, temperature T = 437 °C at the Venus surface) where human living is impossible, should teach us to be cautious with manipulating our own atmosphere.

This textbook will help the reader to follow the growing insight into our universe with all of its objects, and it will induce him to appreciate the work of the early astronomers, and to get more closely involved in the fascinating field of astrophysics. It is primarily not addressing astronomy students, but rather interested physics and chemistry students who had no chance to attend special lectures on astrophysics. The author believes that every intellectual should have at least a basic knowledge of astronomy and astrophysics, because this subject is concerned with fundamental questions about how our life starts and what the future is of our universe, whether there are other civilizations somewhere in the universe and what are the chances to communicate with them.

1.1 History of Astronomy and Astrophysics

Contrary to Nuclear Physics, which is a fairly new branch of physics, Astronomy is one of the oldest natural sciences. Already in ancient times, people have admired the overwhelming view of the sky with thousands of visible stars, and scientists wanted to find out more about stars, their positions and their apparent motions due to the rotation of the earth. In ancient times astronomy and astrology were more closely related and the question, how stars influence our life on Earth and determine the individual fate of a person, was intensely discussed. This superstition is, unfortunately, still not completely abandoned today.

For many centuries the main task of astronomers was the measurement of positions of "fixed stars" at the firmament and the motion of planets and comets. The Babylonians, the Egyptians and the Chinese performed such measurements already several thousand years ago (Fig. 1.1). They drew star maps and planet tables and they gave names to the different stars and planets. They also recognized the importance of star observations for time determinations. When Sirius appeared at the horizon, Egyptian astronomers knew that the Nile flood came, which was essential for fertilizing the fields. They used the star position as time markers but they knew nothing about the properties of stars, except that they could measure their brightness. Also, the distance of the stars from Earth was greatly underestimated.

The Maharadscha Jai Singh built from 1724 to 1734 five observatories in India for measuring star positions, which were essential for the determination of exact timing of annual events in the course of a year (Fig. 1.2).

Also in Europe, people tried to observe stars and measure their positions. A famous site which is believed to be an observatory, several thousand years old, is *Stonehenge* in the southern part of England (Fig. 1.3) where at solstice the sunrise can be observed exactly in the direction defined by the central axis of the stone configuration.

In central and southern America the ancient Mayans were experienced astronomers. They observed not only the stars but followed the path of planets and related them to events on Earth. For example, the Mayan leaders used the

Fig. 1.2 Astronomical observatory Jantar Mantar in New Dehli, India (https://en.wikipedia.org/wiki/Jantar_Mantar,_Jaipur)

Fig. 1.3 Stonehenge in Southern England (https://www.bing.com/images/search?q=stonehenge)

position of Venus to decide about wars with its neighbors. There are several observatories (Fig. 1.4) founded some thousand years ago:

About 600 years ago *Ulug Beg*, the grandson of the famous Mongolian emperor *Timur* founded 1420 in Samarkand, Uzbekistan, a star observatory, where he measured star positions with an accuracy that could compete with that of well-known European astronomers.

In Europe the observations of *Tycho-de Brahe* (1546–1601) (Fig. 1.5b) have to be mentioned, who measured with precise instruments, such as Sextants the path of planets, in particular of Mars, with high accuracy.

Much more detailed observations were possible after the telescope had been invented in 1608 by the Dutch optician *Hans Lipperhey*. It was successfully used by Galilei (Fig. 1.5a) who discovered the 4 Galilean moons of Jupiter with his telescope. This observation unsettled the old hypothesis that the planets were fixed on glass shells which moved around the Earth. The argument of Galilei against representatives of the church was: "*The moons, circulating Jupiter, would destroy the glass shell*".

Fig. 1.4 The Mayan Observatory at Chichen Itza in Mexico (https://www.bing.com/images/search)

Johannes Kepler (1571–1630) (Fig. 1.7), used the data of *Tycho de Brahe* (Fig. 1.5b) to identify for the first time the paths of planets as ellipses instead of circles. His and Galileo's measurements supported the heliocentric model of *Nicolaus Copernicus* (Fig. 1.6) which was forbidden by the church, because it contradicted the belief, that the earth as the unique place in the universe was selected by God and must be therefore the center of the universe.

A much more detailed insight into the structure of stars could be achieved by Astro-Physics, where the knowledge from other branches in Physics, such as atomic physics, nuclear physics, thermodynamics and hydrodynamics could be applied to the study of stars, their interior and their development. Stars were no longer eternal static phenomena, but had an interesting life: They are born from dense gas clouds, contracting into hot stars, which undergo a wild juvenile life but finally reach an equilibrium state which lasts for millions to billions of years Their energy supply comes from nuclear fusion in their central part, and they expire when this energy source is exhausted, ending as **White Dwarfs, Neutron Stars** or **Black Holes**, depending on their mass.

The great success, achieved in Astrophysics during recent years is due to several facts:

(1) The experimental observation techniques have been greatly improved during the twentieth century. Huge parabolic mirrors were constructed and very sensitive detectors were developed.
(2) Satellites outside our atmosphere allow the detection of radiation over large spectral ranges on both sides of the visible spectrum, i.e. in the infrared, the ultraviolet and in the gamma region. The radiation within these regions is absorbed by our atmosphere and can be therefore not observed with earth-bound telescopes. Observations in these extended spectral ranges increase considerably the wealth of information about stars.
(3) An essential contribution to a more detailed understanding of stars was provided by the results of nuclear and high energy physics and hydrodynamics which allowed quantitative calculations about energy production and energy transport from the central part to the surface of stars.

Also, models about the beginning of our universe (Big Bang theory) are based on results of high energy and particle physics. Therefore particle physics and Astrophysics, which seem at first sight to be quite different fields have many connections and both gain from each other. The same is true for the mutual inspiration of Atomic Physics and Astrophysics.

Fig. 1.5 **a** Galileo Galilei (1564–1642). **b** Tycho de Brahe (1546–1601) (https://www.bing.com/images/search)

Fig. 1.6 Nicolaus Copernicus (14731543) (https://de.wikipedia.org/wiki/Nikolaus_Kopernikus)

1.2 Survey on the Concept of This Textbook

The presentation of Astronomy and Astrophysics starts with the question, which kind of information we can obtain from stars, either on the earth's surface or outside our atmosphere. In order to determine the position of stars one has to introduce definite coordinate systems, which take care of the daily and annual variations of the star positions due to the motion of the earth. This requires the exact measurement of time.

The modern equipment and detection techniques of Astronomy, discussed in Chap. 2 have undergone fascinating improvements during recent years, such as active and passive adaptive optics. We will present modern astronomical telescopes on Earth and on satellites, interferometric techniques, X-ray and radio-telescopes, gravitational wave detectors and particle detectors for measuring cosmic neutrinos and particles in the cosmic rays.

With this knowledge, we can learn in Chap. 3 about measuring techniques and possible error sources such as aberration of light from the stars and its refraction in our atmosphere. The daily and annual parallax, due to the daily rotation of the earth and its annual circulation around the sun has to be taken into account in order to determine the true star positions from their measurements on earth. The fluctuation of the atmospheric density causes a corresponding fluctuation of the refractive index, which leads to a blurring of the star image in the focal plane of a telescope. Here special techniques, such as **adaptive optics** have greatly improved the angular resolution of telescopes.

The next Chap. 4 discusses all information we can get about our solar system, including the 8 planets, numerous smaller bodies such as planetoids, meteorites and comets and, of course, our sun. The apparent features of the sun, such as its total radiation with its spectral distribution, its mass and its distance from the Earth will be presented together with the corresponding measuring techniques,

Of particular interest is the energy production in the sun and its transportation to the sun's surface, because we owe our life to the sun's radiation. We will discuss our present understanding of the nuclear processes, that take place in the central part of the sun and how we could determine them from measurements on Earth.

Fig. 1.7 Johannes Kepler (https://www.bing.com/images/search?q=johannes+kepler)

Chapter 5 discusses the properties of stars and ways to measure them. A particularly difficult problem is the determination of the distance of stars. Therefore a series of different measuring techniques have been developed which depend on the distances of the stars and which cover the range from a few light years for stars in our neighborhood to billions of light years for stars in faraway galaxies.

An important quantity of star measurements is the detected radiation and its intensity, which not only depends on the emission of the star but also on its distance from Earth. The spectral distribution of the star emission gives important information about the temperature of the star's surface and the composition of chemical elements in its atmosphere.

Chapter 6 informs about our present knowledge of the life of stars, how they are born, what happens during their life and why they finally die. Their end-states as **White Dwarfs**, **Neutron stars** or **Black Holes** are still the subject of intense research.

Chapter 7 contains a short discourse about Cosmology. It gives a survey about our universe, its development, and theories about its beginning. The question, how our elements have been produced, how stars and galaxies were formed and what the composition of matter and energy in the universe might be, is still a problem of controversial debate. It seems, that besides the ordinary matter, which can be detected by its radiation, there might be much more "dark matter" and "dark energy", which are, however, presently still a mystery and not available for measurements. This is shortly discussed in Chap. 8.

The last Chap. 9 brings us back to our earth. We will discuss, how our solar system has been formed, what the age of our earth is and how one can find out about it. We can learn much about the development of our own solar system from observations of extra-solar planets, which have progressed in recent years quite impressively.

More detailed information can be found in [1–11].

References

1. https://en.wikipedia.org/wiki/Babylonian_astronomy.
2. J.L.E. Dreyer: History of Astronomy from Thales to Kepler (Dover Books on Astronomy 2003
3. R. Buchheim: The Sky is your Laboratory. Advanced Astronomy Projects for Amateurs (praxis Verlag 2007)
4. K. Kellermann, E.N. Bouton, S. Brandt: Open Skies. (Spri nger Heidelberg 2020).
5. H. Kartunen, P. Kröger, H. Oja, M. Putanen, K.J. Donner: Fundamental Astronomy (Springer, Heidelberg 2017)
6. T Dickinson and A. Dyer: The Backyard Astronomers Guide (Springer, Heidelberg)
7. E.N. Penprase: The Power of Stars (Springer Heidelberg 2018)
8. A. Pannekoek: A history of Astronomy. (Dover Publications New York 2003)
9. Tyler Davison: Astronomy. A beginners Guide to Astronomy (2015)
10. O. Usher, Lars Lindberg Christensen: The Universe through the Eyes of Hubble. (Springer, Heidelberg and ESA, 2014)
11. D. Liverington: Encyclopedia of the History of Astronomy and Astrophysics: (Cambridge University Press 2013)

Basic Principles and Techniques of Experimental Astronomy

2

Contents

2.1 Astronomical Coordinate Systems 7
2.2 Temporal Changes of the Coordinates 11
2.3 Definitions of Astronomical Times 12
2.4 The UTC is Nowadays Worldwide Used 12
2.5 Observation of Stars 12
2.6 Telescopes 14
2.7 Detectors for Gravitational Waves 31
Summary 31
Problems 33
References 33

The observation of celestial bodies (stars, planets, comets, asteroids etc.) gives the following information:

- the location of the object in space. In spherical coordinates, this means the determination of the position vector $\boldsymbol{r} = \{r, \vartheta, \varphi\}$ and the choice of a suitable coordinate system, where the position vector is defined. If the object moves in this coordinate system, its position vector $\boldsymbol{r}(t) = \{r(t), \vartheta(t), \varphi(t)\}$ becomes time-dependent.
- The electromagnetic radiation power L_s, which is emitted by the object at the distance r(t) from the earth. The detector receives the fraction $f = L_S \cdot \Delta\Omega/4\pi$, where $\Delta\Omega = F_D/r^2$ is the solid angle, under which the detector area F_D is seen from the celestial object. (The area F_D is for instance the usable aperture of a telescope). The radiation power $L_S(t)$ can be time-dependent (e.g. for variable stars).
- The spectral distribution of the radiation, which can be measured with wavelength-selecting instruments (e.g. spectrographs, interferometers or spectral filters).
  Since our atmosphere absorbs certain spectral ranges of the incident electromagnetic radiation, (Fig. 2.1), detectors on Satellites above the atmosphere, which are not restricted by absorption of the atmosphere, can give us additional information about those spectral ranges missing on the earth's surface.
- Particle radiation (e^-, e^+, p^+, ν, $\overline{\nu}$) which is emitted from celestial objects (e.g. the sun), and which can be measured with special particle detectors.
- Gravitational waves, which are emitted by accelerated massive bodies (e.g. colliding neutron stars, or black holes). They have been recently detected by sophisticated laser interferometers.
- The physical, chemical and mineralogical analysis of meteorites which have fallen onto the earth.

These are the essential sources of information which allow us to obtain more detailed knowledge about the structure, the development and the dynamics of stars, galaxies and the universe. How this information can be combined to gain a consistent model of our world, will be discussed in the following chapters.

2.1 Astronomical Coordinate Systems

In order to define the angular coordinates (which give the location of an object in the sky), it is appropriate to introduce a spherical coordinate system. There are several possibilities [1].

2.1.1 The Horizontal System

From the viewpoint of an observer on the earth's surface, the horizontal system (Fig. 2.2) is most convenient.

The vertical line through the observation point cuts the celestial sphere in the Zenith and on the opposite side in the Nadir (Fig. 2.3a). The plane perpendicular to this vertical line is the horizontal plane, encircled by the horizon. The meridians in this system are the vertical circles through Zenith and Nadir. The circles of latitude are the horizontal circles, also called azimuthal circles. On the horizontal plane, the cardinal points are north point (0°, in the

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_2

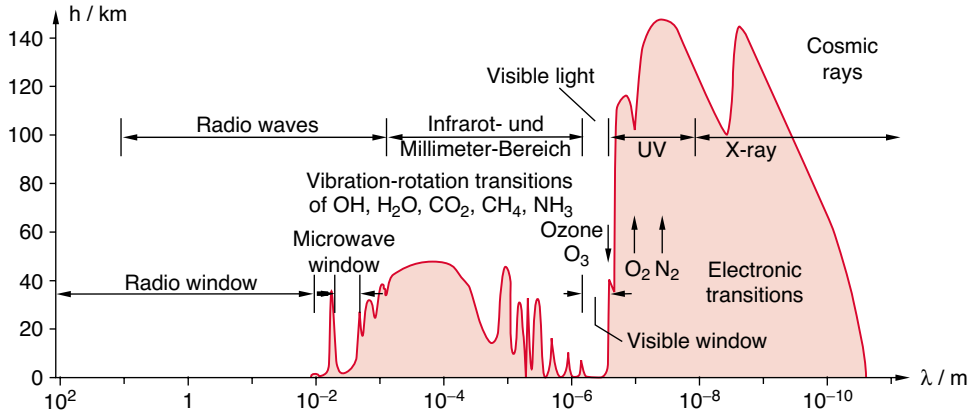


Fig. 2.1 The spectral windows of the earth's atmosphere. The figure shows the heights h above ground, where the radiation intensity I_0 from celestial sources has decreased to I_0/e because of the absorption in the atmosphere

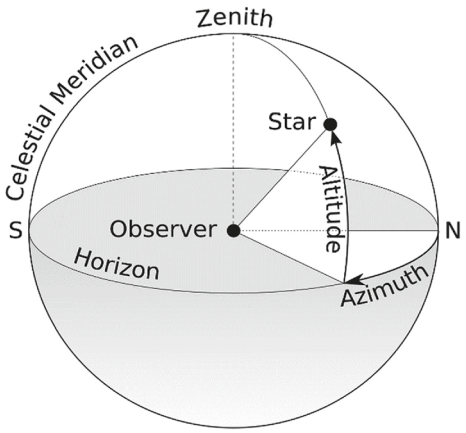


Fig. 2.2 Horizontal coordinate system (Wikipedia)

direction of the earth's north pole), and (in clockwise direction) east (90°), south (180°) and west (270°).

The vertical circle through the north point and the south point is the *celestial meridian*.

The position of a star is characterized by its altitude h and its azimuth A, where h is the angle between horizontal plane and horizontal circle through the star, and A is the angle between meridian and vertical circle through the star, measured from the south point, (chosen as zero point) in the direction west, north, east, south (0–360°). Sometimes instead of the altitude h the zenith distance $z = 90° - h$ is used.

Since the earth rotates about its south-north axis the position of a star in the horizontal system, (which is permanently connected with the rotating earth), changes in time (Fig. 2.3b). For an observer at the latitude φ the star traverses on a circular arc about the earth's axis inclined against the plane of the observer by the angle $(90° - \varphi)$. The length of the circular arc, visible for the observer (*diurnal arc*), depends on the altitude h of a star. For $h = \varphi$ the path of a star above the observation plane is a half circle.

The disadvantage of the horizontal coordinate system is the fact, that the coordinates of a star are time-dependent. Furthermore, the position of the same star is at the same time different for two observers at different locations. Therefore another coordinate system, the equatorial system, is better suited for astronomical observations.

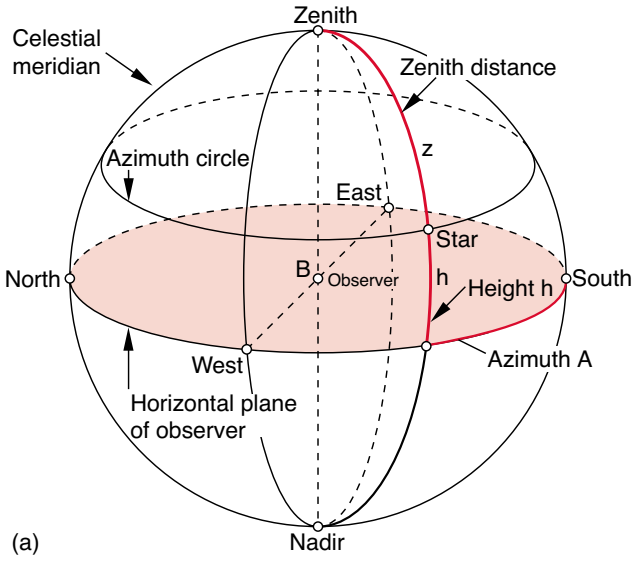
2.1.2 The Equatorial Coordinate System

In the equatorial system, the plane through the earth equator is the reference plane. It cuts the celestial sphere in the celestial equator (Fig. 2.3). The origin of the system is the observation point O in the center of the equatorial plane. The extended earth axis cuts the celestial sphere in the celestial north pole P and in the celestial south pole $\overline{P}$ (Fig. 2.4).

The circles through the two celestial poles are the **hour circles**, the circles parallel to the equator are the parallel circles (Figs. 2.4 and 2.5). The hour circle through Zenith, North- and South points of the horizon and celestial North- and South poles is the **Meridian**. The angular coordinates of a star are defined by the hour circle and the parallel circle through the star.

In the equatorial system, the hour angle t is defined as the angle between the planes through Meridian and the hour circle of the star. The angle between the parallel circle and the equator is the declination δ. It is measured, in analogy to the geographical latitude on earth from $\delta = 0°$ at the equator up to $\delta = 90°$ at the north pole (resp $-90°$ at the south pole).

When the earth rotates, the star essentially remains (aside from precession and nutation of the earth's axis) on its parallel circle i.e. its declination δ remains constant. Its hour circle increases, however, uniformly with time. Since one revolution corresponds to the angle 360° and a sidereal time interval of 24 h, the hour angle changes by 15° per hour. The choice of zero point for the hour angle, which is chosen here as the intersection of the celestial equator with the ecliptic, where the sun appears at the beginning of springtime (vernal equinox at the 21 of March). The ecliptic



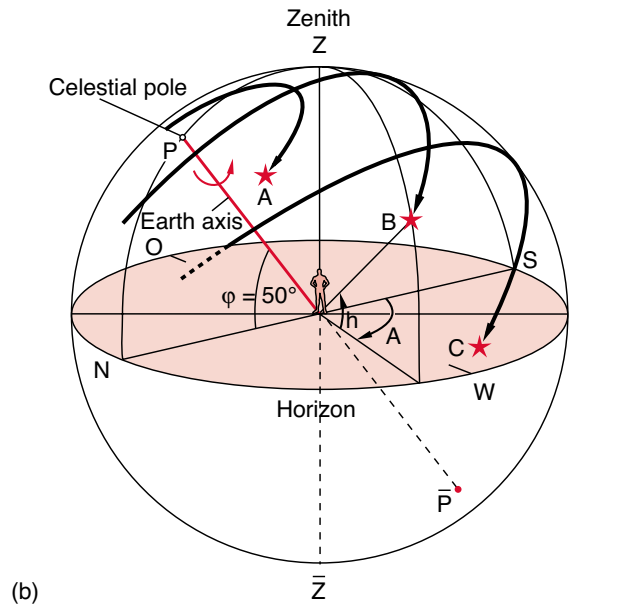


Fig. 2.3 **a** Horizontal system. The observer B sits in the center of the coordinate system. **b** Apparent motion of three stars with different zenith distances around the celestial pole. After Karttunen et al. Fundamental Astronomy [1]

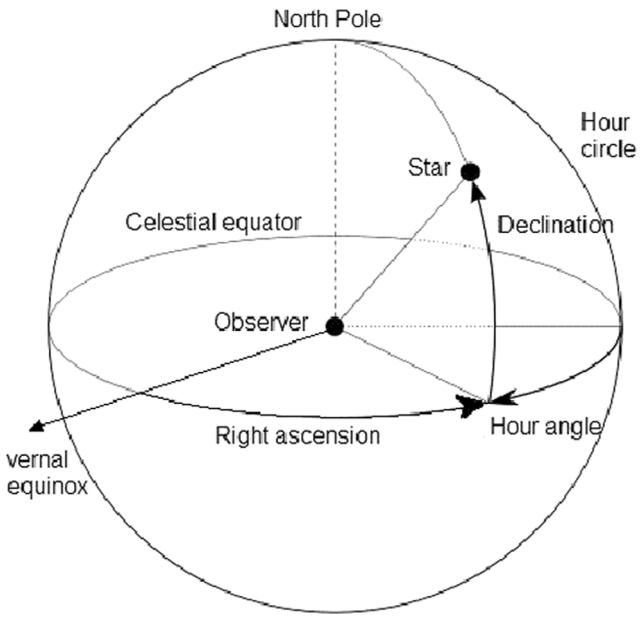


Fig. 2.4 The equatorial coordinate system (Wikipedia)

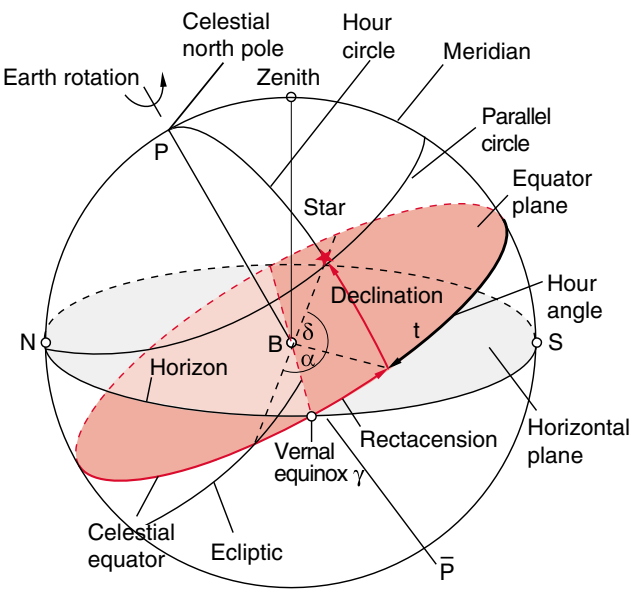


Fig. 2.5 Equator system. B = observer

is the plane of the earth's motion. It intersects the celestial sphere in a circle on which the sun moves as seen from the earth (Fig. 2.6). The angle of the hour circle of a star, measured from the spring point is the **right ascension** (*rectacension*) **α**. It is counted from 0 to 24 h in the direction of the earth's rotation (which means opposite to the apparent motion of the sun) starting from the spring point (equinox). Since a star participates in the apparent daily revolution of the celestial sphere its right ascension remains constant.

In the equatorial coordinate system, the angular coordinates (declination δ and rectacension α) of a star are equal for all observers on earth and at any time (taking aside the precession and nutation of the earth axis). They can be therefore compiled in astronomical tables.

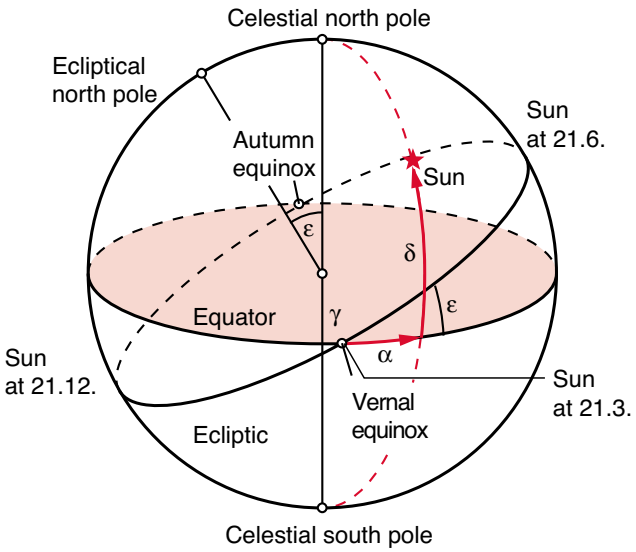


Fig. 2.6 Position of the ecliptic relative to the equator plane. The angle ε gives the inclination of the earth's axis against the normal to the ecliptic

2.1.3 The Ecliptical System

For the description of planets and other objects of our solar system (see Chap. 4) a coordinate system is best suited which uses the ecliptic, i.e. the orbital plane of the earth as the reference plane (Fig. 2.7). The center point in this plane is either the sun (heliocentric system) or the earth (geocentric system).

The ecliptic plane intersects the celestial sphere in a circle which equals the orbit of the sun on its apparent motion, when it passes the equator plane in the vernal equinox on the 21st. of March, moving from the southern to the northern hemisphere (Fig. 2.5). Declination and right ascension of the vernal equinox are both zero in the ecliptical system.

The position of a celestial object in the geocentric ecliptical coordinate system is defined by the angular distance

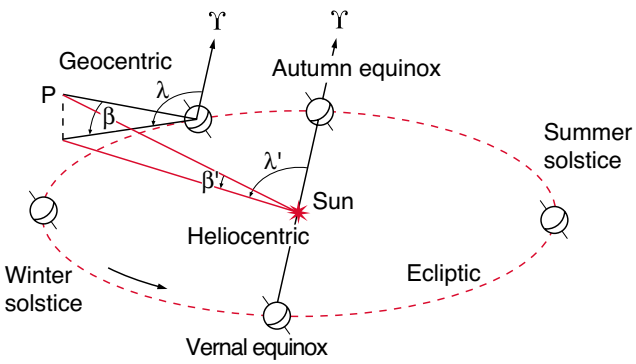


Fig. 2.7 Ecliptical system with the heliocentric coordinates ß′ and λ′ and the geocentric coordinates ß and λ. After Karttunen et al. [1]

ß from the ecliptic plane (*ecliptical latitude*) and the angle λ of the ecliptical length (*ecliptical longitude*) which is counted from the vernal equinox anti-clockwise. In the heliocentric ecliptical system, the latitude is ß' and the longitude is l. The origin of the ecliptical system is generally the center of the sun (heliocentric system) or the center of the earth (geocentrical system).

While the position of objects at far distances does not differ much for the two different choices of the origin, it is, of course, essential for the position of objects in our solar system.

2.1.4 The Galactic Coordinate System

The coordinate systems discussed so far are either connected with the earth (geocentrical systems, e.g. the equatorial system or the geocentric ecliptical system), or with the sun (heliocentric ecliptical system). Although the equator system takes into account the daily rotation of the earth and its annual revolution around the sun, there are long term variations (such as the precession and nutation of the earth's axis) which demand corrections of the angular positions of stars.

The differential revolution of our solar system around the center of our galaxy results in further long term changes of the star coordinates (see Sect. 7.5).

For measurements of the spatial distribution of the stars in our galaxy and the definition of their positions, the galactic coordinate system is better suited (Fig. 2.8). Its reference plane is the galactic equator plane, which nearly coincides with the central plane of the milky way band seen in dark nights on the sky. This plane is inclined against the ecliptic by 57°.

The vertical line through the center of the galactic plane extends to the galactic north pole above the plane and to the galactic south pole below the plane. The great circles through the poles are the meridians, and the circles parallel to the galactic equator are the circles of latitude.

The galactic coordinates are the galactic longitude l and the galactic latitude b. They are defined as follows: The galactic longitude l is the angle between the line sun-galactic center and the projection of the circle of longitude through the star onto the galactic plane, measured anti-clockwise from the line sun-galactic center (Fig. 2.8b). The coordinate $l = 0$ gives therefore the direction sun-galactic center. The galactic latitude b is the angle between the galactic plane and the line from the sun to the star. It is positive ($b > 0$) for stars above the galactic plane and negative below the plane.

In order to compare the galactic coordinates with the coordinates in the equatorial system the International Astronomical Union IAU has defined the following relations:

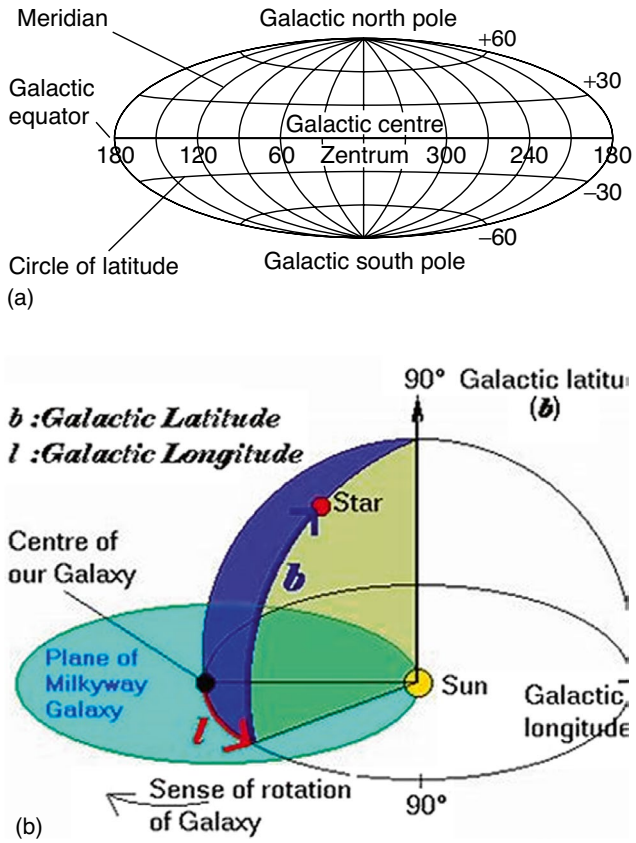


Fig. 2.8 The galactic coordinate system. **a** Survey. **b** The galactic coordinates (l = galactic longitude, b = galactic lattitude). (College Inleiding Sterrenkunde (rug.nl))

The galactic northpole has the right ascension $\alpha = 12^h45^m$ and the declination $\delta = 27.4°$ and the galactic center has the coordinates $\alpha = 17^h45^m37.224^s$, $\delta = -28°56'10.23''$.

Recent more accurate measurements have shown that the galactic center, located in the constellation Sagittarius, slightly deviates from these defined coordinates (see Table 2.1) [3].

2.2 Temporal Changes of the Coordinates

As has been discussed in Vol. 1, Sect. 5.8 the direction of the earth's axis changes in time. Caused by torques on the earth as oblate rotational ellipsoid, exerted by the sun, the moon and other planets, the earth's axis precesses within 26.000 years on a cone with opening angle $2 \times 23°26'$ around an axis vertical to the ecliptic. This precession modulates right ascension (rectacension) α and declination δ of a star. As has been derived in [4] the annual changes are

$$\Delta\alpha = m + n \cdot \sin\alpha \tan\delta, \tag{2.1a}$$

$$\Delta\delta = n \cdot \cos\alpha, \tag{2.1b}$$

where $m \approx 3.07\text{s/a}$ and $n \approx 1.33\text{s/a} = 20.0''/\text{a}$ are the **precession constants**, which can be found in astronomical tables. (1 s is one time unit, whereas $1''$ is a second of arc.)

In addition to the precession, another uniformly periodic perturbation of the precession occurs which is caused by the moon. It has a period of 18.6 years and is called an Astronomy *Nutation*. It is induced by the periodic change of the gravitational force between earth and moon. The orbital plane of the moon is inclined against the ecliptic by 5°. Due to the gravitational attraction of the moon by the sun, the normal orbital plane of the moon executes a precession around the pole of the ecliptic with a period of 18.6 years. This causes a periodic change of the direction of the earth's axis and with it a corresponding change of the angular velocity of the earth.

In addition to these periodic changes, there are some irregular fluctuations in the direction of the earth's axis, caused by changes in the mass distribution inside the earth (drifts of glaciers, earthquakes, volcanic eruptions, etc.).

This illustrates that in the equatorial system, the angular coordinates α and δ are by no means completely constant. Therefore one has to add the annual time they are referred to. This time is called the Epoch. It is added behind the specification of the angles α and δ in brackets [6].

Note: Usual epochs are 1900 (January 1st 1900, 12 h universal time UT) or 1950 or 2000.

Example

The equatorial coordinates of the giant galaxy NGC 12.75 in the Perseus Galaxy cluster (Fig. 7.24) are ◀

$$\alpha(2000.0) = 3^h19^m48.^s159,$$
$$\delta(2000.0) = 41°30'42.''10.$$

Note: that it has become custom in Astronomy, to place the angle and time units above the decimal point, Numerical

Table 2.1 Equatorial coordinates J 2000.0 of galactic reference points https://en.wikipedia.org/wiki/Galactic_coordinate_system-cite_note-Blaauwetal-1 (Wikipedia)

	Right ascension	Declination	Constellation
North pole +90° latitude	$12^h51.4^m$	+27.13°	Coma Berenices (near 31 Com)
South pole −90° latitude	$0^h51.4^m$	−27.13°	Sculptor (near NGC 288)
Center 0° longitude	$17^h45.6^m$	−28.94°	Sagittarius (in Sagittarius A)
Anticenter 180° longitude	$5^h45.6^m$	+28.94°	Auriga (near HIP 27,088)

values of the luminosity are written the same way (see Sect. 2.8).

2.3 Definitions of Astronomical Times

Since the measured star positions change with time, an accurate time measurement is essential [2].

The unit of time is the second, which is realized by several cesium atomic clocks in selected labs around the world (see Vol. 1, Sect. 1.6). In Astronomy the international atomic time (TAI) is defined as the average of the time, measured by several atomic clocks (Braunschweig, Germany, Boulder, Colorado USA, London, Paris). In 1972 a coordinated universal time (UTC) was introduced which is realized by broadcasting time signals which define the accurate time for every place on earth. This UTC is equal to the TAI plus the time zone difference of the particular location.

The basis of the astronomical time systems is the mean apparent time of Greenwich (see Vol. 1, Sect. 1.6.3), which takes into account the precession and nutation of the earth's axis. It is called the **universal time UT1**. The time difference

$$\Delta\mathrm{UT} = \mathrm{UT1} - \mathrm{UTC}$$

between universal time UT1 and coordinated universal time UTC considers the variations of the earth's rotation. The actual values of ΔUT are regularly published by the institutes for time measurements.

2.4 The UTC is Nowadays Worldwide Used

In order to adapt the UTC-time as accurately as possible to the UT1 time derived from the earth's rotation, deviations due to the changing rotation period, are compensated by **leap seconds**. This ensures that the difference between UTC-time and astronomically defined UT1 never exceeds 1 s. Since the earth's rotation permanently slows slightly down (due to friction), since 1972 in the mid of a year or at its end, a leap second is inserted, if necessary.

Besides the second, larger time units are used in astronomy:

(1) The solar year (tropical year) is defined as the time interval between two successive transits of the sun through the vernal equinox. (remember: This is the intersection between the ecliptic and celestial equator through which the sun transits on the 21st of March).
(2) The true solar day is the time interval between two successive lower meridian transits of the sun (midnight). Since the length of the true solar day changes during the year (the earth's rotation is not uniform and the velocity of the earth on its revolution around the sun shows periodical fluctuations) the mean solar day is introduced. This is the time interval between two successive lower mean solar days. It is the average of all 365 days during the year. It has exactly 24 h. The astronomical definition is.
The mean solar day is the time interval between two lower meridian transits of a fictive sun, which performs (seen from the earth) with uniform velocity (averaged over the year) its passage along the celestial equator. The solar year has 365.25 mean solar days.
(3) The *siderial day* is defined as the time interval between two successive transits of the same star through the meridian. Since the earth runs once per year around the sun (this gives one more full rotation as seen from the star) there is one more siderial day per year. The year has therefore 366.25 siderial days. A siderial day is about 4 min shorter than a *synodial day* (see Sect. 3.1.4).

In order to facilitate the conversion of calendar dates and the detection of comet periods, in 1581 the *Julian datum* was introduced. It gives the number of mean solar days since the 1st of January in the year −4712 (4713 after Christ) at 12 o'clock universal time UT. To avoid too large numbers the modified Julian datum MID was introduced where the zero point was set to the 17th of November 1858 at 0 o'clock universal time. The conversion is $\mathrm{MJD} = \mathrm{JD} - 2.400.000,5$.

Example

The 1st of January 2000 at 12 o'clock corresponds to the Julain date $\mathrm{JD} = 2.451.545,0$ and $\mathrm{MJD} = 51.544,5$. ◀

2.5 Observation of Stars

In order to get a better orientation with respect to star positions and to find a star with the numerical values of right ascension α and declination δ obtained from astronomical tables, it is helpful to realize the apparent motion of the stars for different observation points on earth.

For an observer at the north pole, the celestial pole coincides with the zenith, The apparent orbits of all stars are circles around the north pole (Fig. 2.9) which are traversed clockwise. For an observer at the south pole, the stars traverse in circles anti-clockwise. The radius of these circles depends on the declination δ of a star. For $\delta = 90°$ the radius is zero (Fig. 2.10).

For an observer at the equator, the celestial north pole P=N and south pole $\overline{P} = S$ are in opposite directions on

Fig. 2.9 Exposure of the sky with a fixed telescope directed to the north (exposure time 3 h), which shows the apparent circular orbits of stars around the south-north axis due to the rotation of the earth (from Gondolatsch: Astronomie, with kind permission of Klett Verlag)

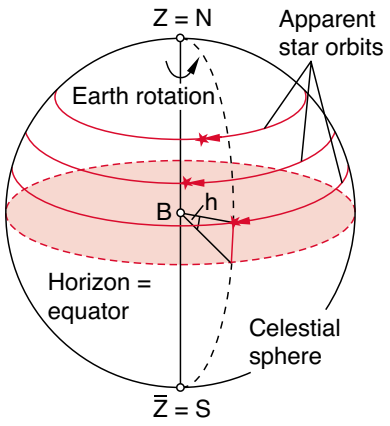


Fig. 2.10 The celestial sphere and the apparent motion of stars for an observer at the north pole

the horizon. The orbits of visible stars are semi-circles around the earth's axis N-S (Fig. 2.11). Their radius is again dependent on the declination δ of the star. It is zero for $\delta = 90°$ and maximum for $\delta = 0°$. The stars traverse the circles clockwise, when observed from the north point N and anti-clockwise when observed from S.

For an observer at the arbitrary geographical latitude, φ the angle between zenith and celestial north pole is $(90° - \varphi)$. The pole distance is then $h = \varphi$ (Fig. 2.12).

A star with the declination $\delta = 0$ traverses during the rotation of the earth at the celestial equator (Fig. 2.13). It

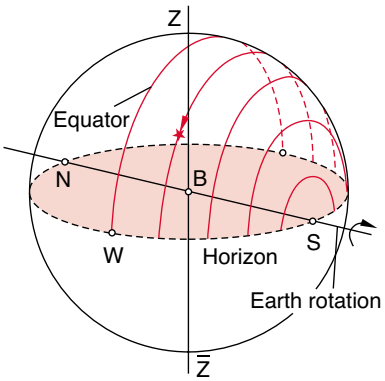


Fig. 2.11 Apparent motion of stars for an observer at the earth's equator

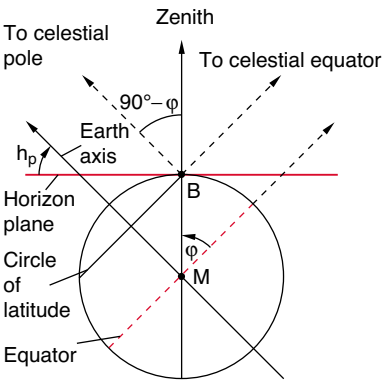


Fig. 2.12 Observation conditions for an observer at the latitude φ

rises in the east point O of the horizontal plane, reaches its highest elevation (culmination) with $h = 90° - \varphi$ in the south point S and sets in the westpoint W of the horizontal plane. It stays for equal times above the horizon (diurnal arc) and below the horizon (night arc).

For $\delta > 0$ rise point and set point both are shifted to the north. The diurnal arc is then for an observer on the northern hemisphere longer than the night arc. For $\delta < 0$ rise point and set point shift to the south, which implies that the diurnal arc becomes shorter and the night arc longer (Fig. 2.13).
Note: The diurnal arc is that part of the star orbit when the star can be seen. Since stars can be generally seen only during night time, between sun set and sun rise only those stars can be seen with a diurnal arc which is traversed during the night arc of the sun.

The heights h of the upper and lower culmination of a star with the declination δ for an observer at the geographical latitude is

$$h_{\text{upper}} = \delta + (90° - \varphi) \tag{2.2a}$$

$$h_{\text{lower}} = \delta - (90° - \varphi) \tag{2.2b}$$

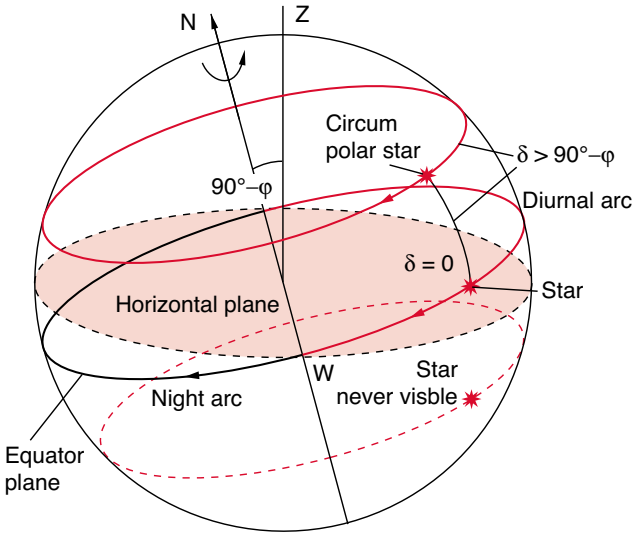


Fig. 2.13 Diurnal arc and night arc of a star with declination δ for an observer at the latitude φ

For an observer at the geographical latitude, φ a star never sets, if $h_{\text{lower}} > 0°$ i.e. $\delta > (90° - \varphi)$. Such stars are called **circum polar** (Fig. 2.13).

One example is the constellation of the great bear (Ursa Major) with the declination extending from $\delta = 30°$ to $\delta = 73°$. It is therefore circum polar for observers at latitudes $\varphi > 90° - \delta \approx 60°$.

A star can never be seen by an observer at the latitude φ if $h_{\text{upper}} < 0°$, i.e. if its declination $\delta < \varphi - 90°$. One example is *Canopus* ($\delta = -52.40°$) which is the brightest star in the constellation Carina (α Carina) and which cannot be seen in our latitude $\varphi = 50°$.

Our sun traverses on the great circle which is the intersection of the ecliptic (plane of the earth's orbit) with the celestial sphere In Fig. 2.14 the orbits of the sun are shown for 4 selected days in the year. It traverses on these orbits from east to west. The sunset points on the horizon which lie around the west point W and cover during the year the range between the locations a and b.

Because of the inclination of the earth's axis against the orbital plane, the angle between the ecliptic and equator is 23.5°. The great circles of ecliptic and celestial equator intersect at two points: The vernal equinox (star constellation Aries) which the sun passes on March 21st from south to north and the autumn point (in the constellation of Libra) where the sun passes at September 23rd from north to south.

At the summer solstice (22nd of June) the sun has its largest declination $\delta = +23.5°$ and at the winter solstice its lowest declination $\delta = -23.5°$ (Fig. 2.15).

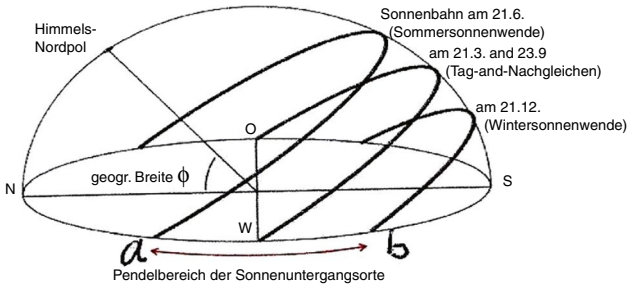


Fig. 2.14 Apparent orbits of the sun, seen from the earth at 4 different dates in the year

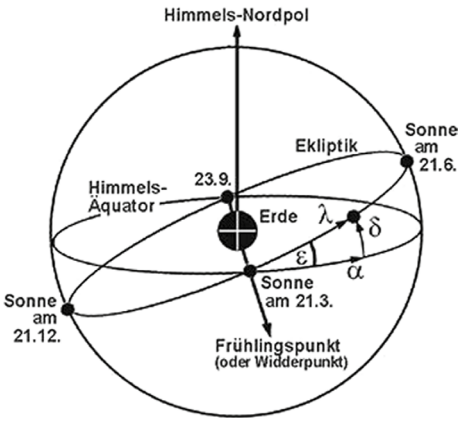


Fig. 2.15 The orbital plane of the sun is the ecliptic

2.6 Telescopes

With the naked eye, one can see from a dark point on earth about 3000 stars. However, alone in our galaxy (the Milky Way) there are about 10^{11} stars, which can be only seen with telescopes, because they are too weak for the naked eye. Besides our own galaxy, there are many millions of other galaxies in the universe, which cannot be seen (except the direct neighbor galaxy) without telescopes [3, 4].

The word "telescope" comes from the Greek (tele = far, scopein = to look) because the first telescopes had been used for looking into far distances on earth, e.g. to detect approaching enemies already at far distances.

2.6.1 Light Gathering Power of Telescopes

One of the main advantages of telescopes is their larger light gathering power of the radiation from celestial objects. Assume d as the diameter of the eye pupil and D as the diameter of the telescope aperture. The ratio of the received

radiation powers from a star (which is assumed as point like radiation source) is then

$$G^2 = \frac{\pi D^2}{\pi d^2} = (D/d)^2. \tag{2.3}$$

Example

$d = 5$ mm, $D = 5$ m $\Rightarrow G^2 = 10^6$. ◀

The telescope receives 1 million times more radiation power than the naked eye.

The different types of telescopes can be divided into lens telescopes (refractors, Fig. 2.16a) and mirror telescopes (reflectors Fig. 2.16b). The first telescopes used by Galilei (Fig. 2.17) and Kepler (Fig. 1.6), were lens telescopes.

The largest refractors have lens diameters of about 1 m (Fig. 2.18).

It is technically difficult to produce bubble-free lenses with diameters $D > 1$ m. Such large lenses have the further disadvantage, that they can be only fixed at their verge and therefore are deformed by the gravitational force of their weight.

Therefore nowadays all large telescopes are mirror telescopes [5]. They have the further advantage that they avoid chromatic aberration and in the form of parabolic mirrors also the spherical aberration. (Vol. 2, Sect. 9.3). Some large refractors, built in 1887, (Fig. 2.18) had been still in use some years ago for special tasks. For example, the refractor of the Yerks observatory had been used for the most accurate laser-distance measurements between earth and moon. [7], Most of these lens telescopes only serve today for special tasks and also as museum objects.

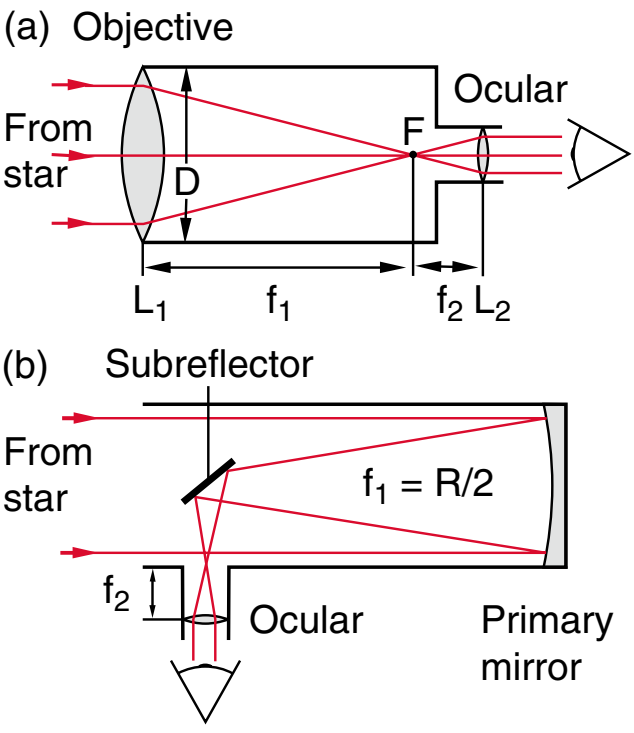


Fig. 2.16 Schematic comparison of a lens telescope (refractor) and a mirror telescope (reflector). The focal length of the parabolic mirror is f = R/2, where R is the radius of curvature of a spherical mirror that touches the parabolic mirror at the center with the same curvature

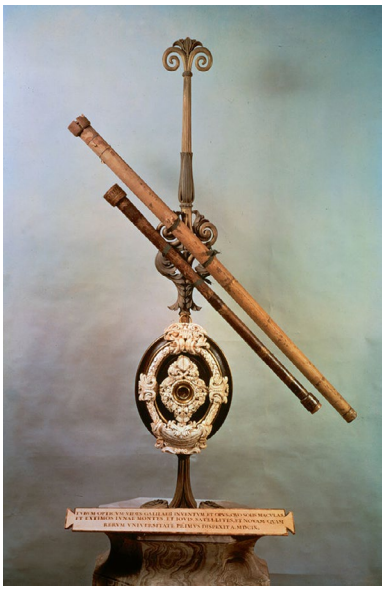


Fig. 2.17 Original lens telescope of Galilei (https://www.mpg.de/7913340/Galileo-Galilei-telescope)

Most observations of stars or galaxies are not performed visually but with sensitive detectors, such as CCD arrays or photoplates, because they are more sensitive than the human eye and they can integrate the received radiation power over a longer time interval. This improves considerably the lower detection limit and opens the way for detecting galaxies at far distances which had been generated at early times of the universe [6].

In order to direct the telescope during the exposure time always exactly towards the same point on the sky, it must compensate for the rotation of the earth by a corresponding opposite rotation (Sect. 2.6.5).

In Fig. 2.19 three different optical arrangements of reflector telescopes are shown: In the **Cassegrain arrangement** (Fig. 2.19a) the incident light hits the large primary parabolic mirror which collects the light onto the small flat subreflector which focuses the light through a small hole in the primary mirror onto the detector. A modification of the Cassegrain arrangement is the Rithexy-Cretiene telescope, which uses two hyperbolic mirrors instead of the parabolic and flat mirror. It is advantageous for wide field observations.

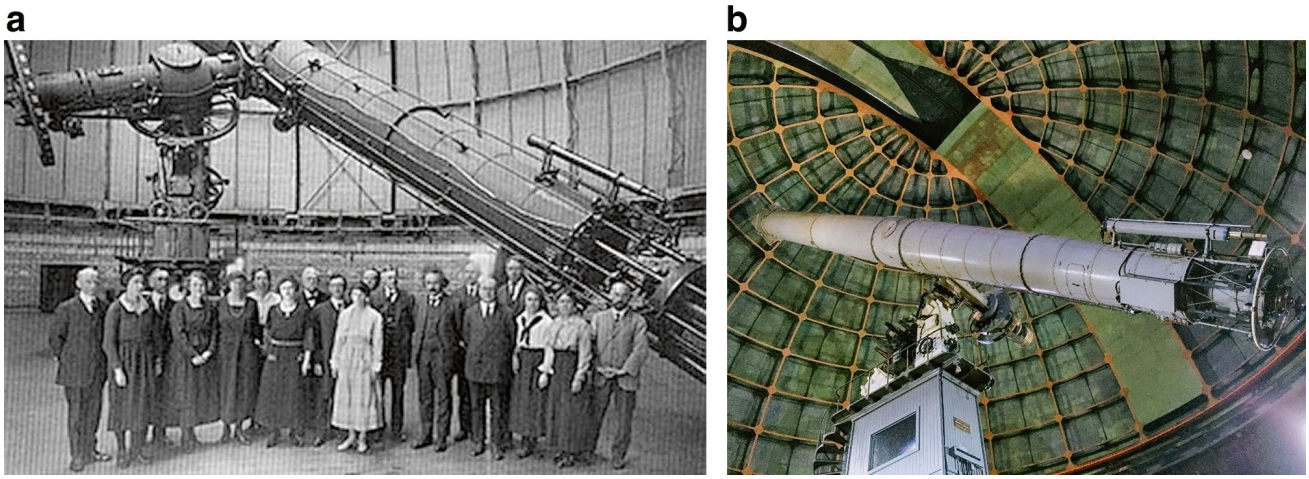


Fig. 2.18 Large refractors. **a** Yerkes observatory (https://en.wikipedia.org/wiki/Yerkes_Observatory). **b** Lick observatory (https://www.bing.com/images/search)

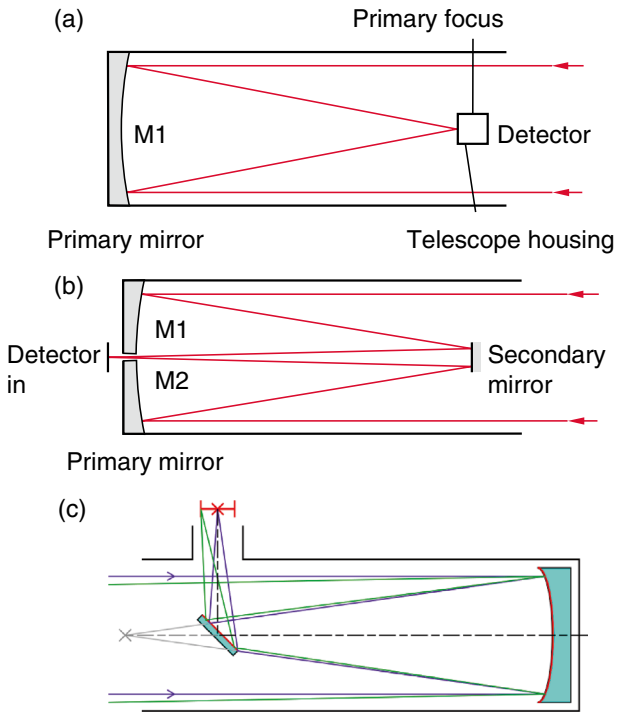


Fig. 2.19 Some realizations of refractors: **a** Cassegrain arrangement. **b** Telescope with only one mirror, where the detector sits in the focal plane of the primary mirror. **c** Newton reflector telescope design

In a system with only one mirror (Fig. 2.19b), the primary mirror focuses the incident light directly onto the detector.

A special design, used by Isaac Newton is the Newton reflector, (Fig. 2.19c), where the flat subreflector focuses the light under an angle of 90° sidewards onto the detector. This has the advantage, that the detector does not block part of the incident light.

2.6.2 Angular Magnification of Telescopes

For observations of extended objects (sun, moon, galaxies, nebula) the magnification of the telescope (angular magnification) is important. As has been discussed in Vol. 2, Sect. 11.2.3, the angular magnification of a lens telescope is equal to the ratio

$$V = f_0/f_e \tag{2.4}$$

of the focal lengths f_0 of the objective and f_e of the ocular. As can be seen in Fig. 2.20 is

$$\begin{aligned} &\tan\beta = b/f_e; \quad \tan\alpha = b/f_0 \\ &\Rightarrow \tan\alpha/\tan\beta \approx \alpha/\beta = f_\beta/f_0 \end{aligned} \tag{2.5}$$

For mirror telescopes, f_0 is the focal length of the primary mirror and f_e is that of the ocular resp. the secondary mirror.

In order to reach a large angular magnification (which is essential for measurements of parallaxes) the focal length f_0 of the primary mirror should be as large as possible, and that of the secondary mirror as small as possible.

The most important properties of a telescope are, however, the angular resolution and the light gathering power. Both quantities depend on the diameter D of the limiting aperture, which generally is the diameter of the objective lens for refractors or of the primary mirror for reflectors.

2.6.3 Different Types of Telescopes

In the Cassegrain reflector (Fig. 2.19a) the incident light falls onto the primary parabolic mirror which reflects and focuses it either directly onto the detector (b) or onto the secondary mirror (a) which focuses it further through a

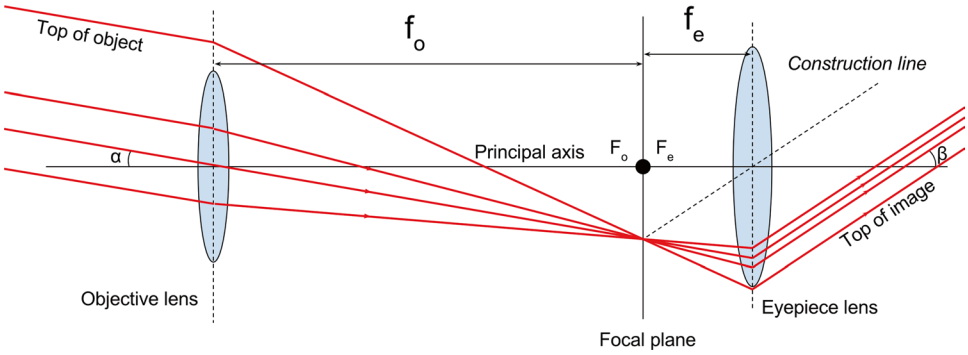


Fig. 2.20 Angular magnification of the Kepler telescope (Mr. Too Good Physics)

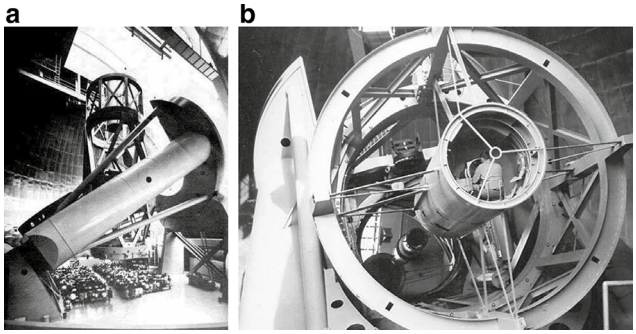


Fig. 2.21 **a** Hale Reflector. **b** Detector cabin in the large Hale reflector on Mount Palomar (Note the man sitting in the cabin) (Palomar Observatory)

small hole in the primary mirror onto the detector. In some arrangements, the secondary mirror is tilted and focuses the light in the direction perpendicular to the incident direction onto the detector (Fig. 2.19c). In very large mirror reflectors the detector cabin with diameter d is located just before the primary mirror (Fig. 2.21b). It blocks, of course, part of the light, but the blocked fraction is very small.

Example

Mirror diameter $D = 5$ m, detector cabin diameter $d = 1$ m $\Rightarrow$ the fraction of incident blocked light is $\eta = (d/D)^2 = 0.04$ i.e. only 4% of the incident light is blocked by the detector cabin. ◀

The large cylindric rod in Fig. 2.21a is just the counterweight of the telescope mounting, which guarantees, that the telescope with many tons of weight is balanced and can be directed into any possible direction with a minimum of torque. The auditorium below the telescope elucidates the large dimensions of the telescope. In Fig. 2.21b the detector cabin with an astronomer looking at the detector illustrates the relative sizes of the detector cabin and mirror.

Another example of a reflecting telescope with a different mount is the 2.2 m reflector at the Calar Alto in Spain (Fig. 2.22).

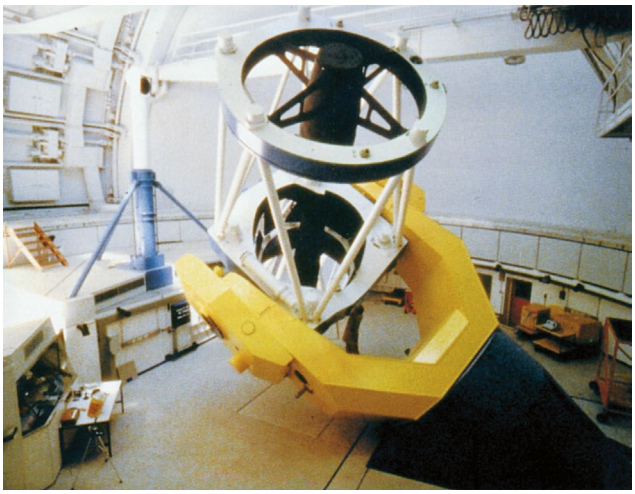


Fig. 2.22 The 2.2 m telescope in its dome on the Calar Alto in Spain. To the left is the blue-white periscope of the Coudé spectrograph (which is located in the floor below). While the focal length of the Cassegrain telescope is 17 m, that of the Coudé is 88 m. The mass of all movable parts in the total amount is 72 tons, while the mass of the telescope is only 15 tons (with kind permission of the Max-Planck Institute for Astronomy in Heidelberg)

For spectroscopic investigations of the spectra of stars, a spectrograph is placed behind the telescope. It should be constructed in such a way that the light always reaches the entrance slit of the spectrograph even when the telescope is redirected to another star. The ingeneous idea of a kinked (French: *coudé*) light path meets these requirements (Fig. 2.23): The light path proceeds from the secondary mirror along the optical axis of the telescope and is reflected by the mirror S along the hour axis onto the entrance slit of the spectrograph. This light path which can be realized by one or several mirrors (Fig. 2.23b) is independent of the rotation of the telescope around the hour axis. The mirror S is adjusted by a computerized motor when the telescope is redirected towards another star [8].

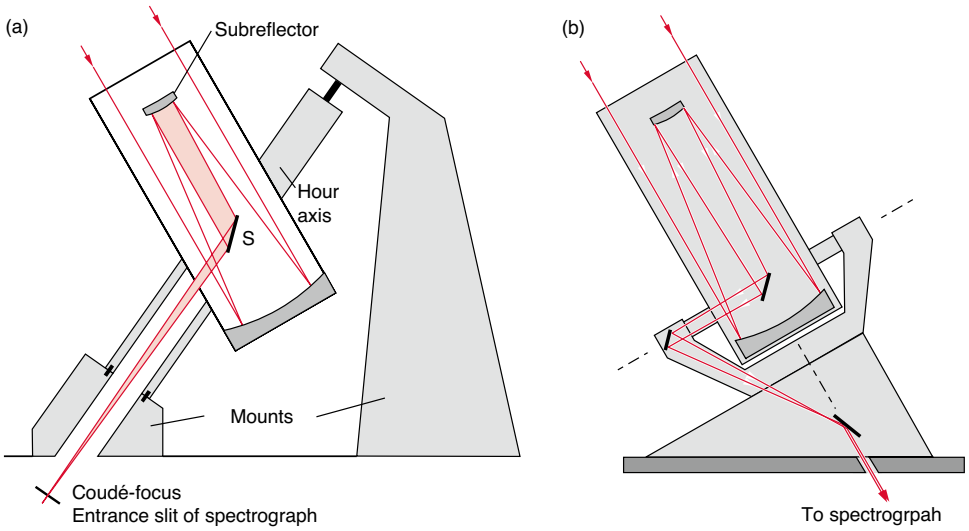


Fig. 2.23 Coudé-beam path **a** with synchronously turned deflection mirror. **b** With three fixed deflection mirrors

Remark There are more complex arrangements with 3 or more mirrors, where the tracking of the mirror S is no longer necessary.

Not only lenses but also mirrors have image errors. They appear e.g. as Koma, if the incident light is inclined against the telescope axis. These inclined light rays are no longer focused into a point, but the image of a point-like star is an extended spot which looks like a comet tail. Such image errors limit the useful field of view of the telescope. For example, the usable field of view of the 5 m Palomar telescope is only $4' \times 4'$. This corresponds only to about 1/8 of the sun's diameter, i.e. 1/64 of the sun's surface.

For survey scans of larger areas on the sky, a larger field of view is desirable. Therefore the optician *Bernhard Waldemar Schmidt* (1879–1935) at the observatory in Hamburg, Germany, developed a special telescope design called the **Schmidt Telescope** (Fig. 2.24). Instead of a parabolic primary mirror, a spherical mirror is used. The spherical aberration is compensated by a thin aspherical glass plate, located in the center of curvature of the spherical mirror [9].

With this Schmidt telescope a large field of view (several degrees) can be realized (the first telescope, built by Schmidt reached 16°!). Such telescopes with a large field of view are in particular useful for the observation of extended nebulae such as the Orion nebula (Fig. 5.13) and nearby galaxies or star clusters.

One disadvantage of the Schmidt telescope is, besides its large tube length (which is L = R instead of L = R/2 as in the Casegrain telescope) the spherically curved focal plane The photoplates (or the CCD detector arrays) must be therefore slightly curved.

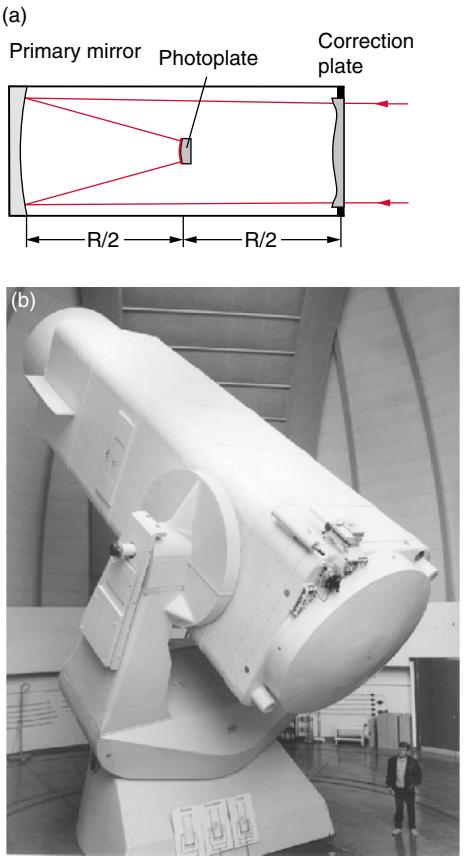


Fig. 2.24 **a** Design of a Schmidt telescope. **b** The largest optical Schmidt telescope is at the observatory in Tautenburg close to Jena (with kind permission of the Thüringer Landes-Sternwarte Tautenburg)

2.6.4 Angular Resolution of Telescopes

Essential for the spatial resolution of the images of stars with a small angular difference in the detector plane is the angular resolution power of the telescope (see Vol. 2, Sect. 11.3). It is in principle limited by diffraction. With a useful diameter D of the telescope, the lower angular limit for the still resolvable images of two stars is

$$\sin \alpha_{\min} = 1.22 \cdot \frac{\lambda}{D} \tag{2.6}$$

This theoretical lower limit can be indeed reached for a telescope outside our atmosphere (e.g. Hubble space telescope) and after the elimination of all relevant imaging errors.

For telescopes on earth the air turbulence is the essential limiting factor for the angular resolution (seeing).

Example

$\lambda = 0.5\,\mu\text{m}, D = 3\text{ m} \Rightarrow \alpha_{\min} = 0''.04$. The air turbulence deteriorates, however, the resolution to about 0.5–1″ which is more than 10 times worse than the diffraction limit. ◀

For comparison: The angular resolution of the human eye is about 1′, which is 60 times worse than the telescope resolution, limited by air turbulence.

Note: The diffraction limit becomes equal to the air turbulence limit for telescopes with $D = 0.15$ m. Larger telescopes on earth bring therefore (without adaptive optics) no advantage regarding the angular resolution, but of course with respect to the light gathering power and sensitivity.

The perturbing influence of the atmosphere can be largely eliminated by the technique of **Adaptive Optics** (Sect. 2.6.5). Another technique, often used before the adaptive optics was introduced, is the **Speckle Interferometry** [10], which allows one to nearly reach the diffraction limit. It works as follows.

One makes a very short (<50 ms) exposure of a star. During this short time interval, the refractive index of the air does not change much and therefore the image is diffraction limited. Sometime later a second exposure is made. Again a diffraction limited image is produced but it appears at another location on the detector plane because the refractive index of the atmosphere will have changed meanwhile. In this way, many short exposures are made and a special computer program analyzes the random distribution of the many diffraction limited star images (Fig. 2.26) and constructs the final image of the star, which is again nearly diffraction limited. The disadvantage of this method is the fact that due to the short exposure time, the single diffraction limited star images are very faint and therefore the technique is restricted to brighter stars.

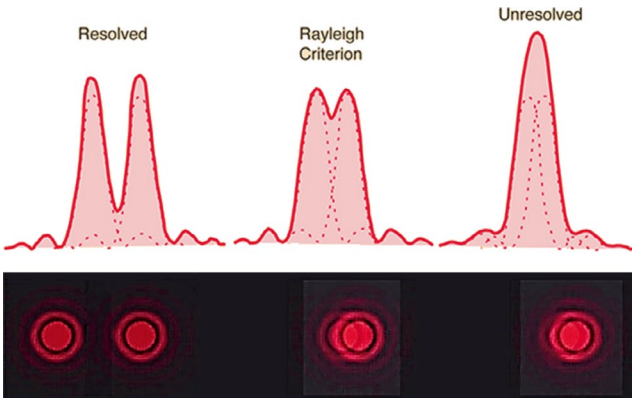


Fig. 2.25 Illustration of diffraction limited resolution of the images of two point sources (http://hyperphysics.phy-astr.gsu.edu/hbase/phy-opt/Raylei.html)

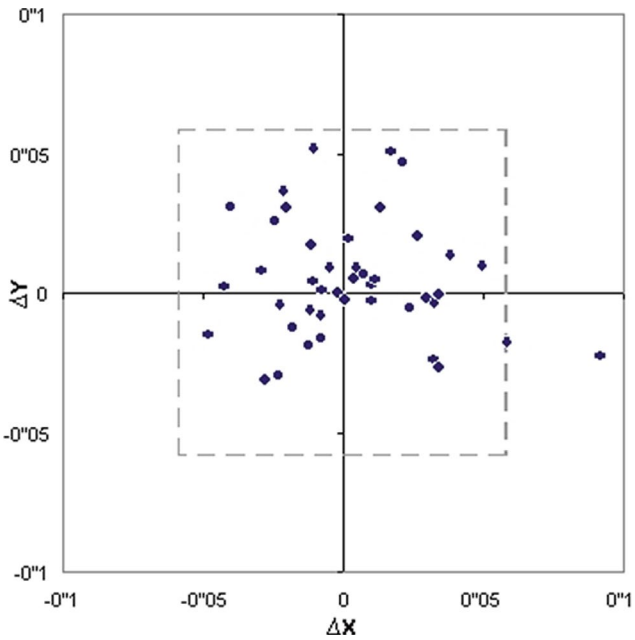


Fig. 2.26 Successive short time exposures of images of the same star (International Meeting of Double Star observations. Barcelona, Spain 2010)

2.6.5 Active Optics

For large telescopes, the weight of the primary mirror represents a problem. In order to reduce the mirror weight, one reduces the thickness of the mirror. However, the resulting disadvantage is the deflection of the mirror, because it can be only held at its edges. This deflection

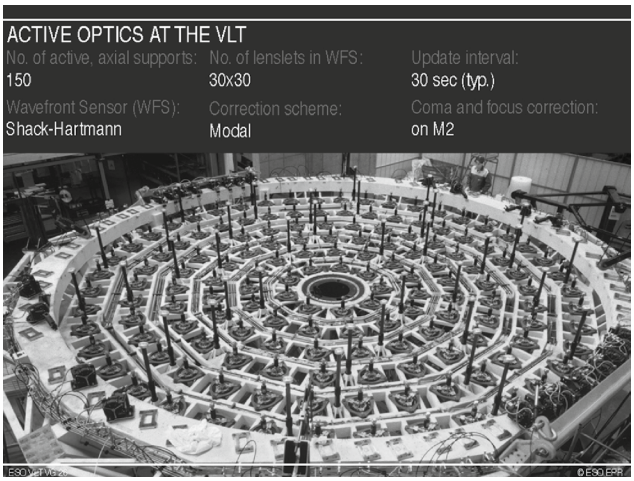


Fig. 2.27 Backside of the 10 m mirror of the VLT on the paranal. The structures are the piezo-elements for the corrections of the mirror surface

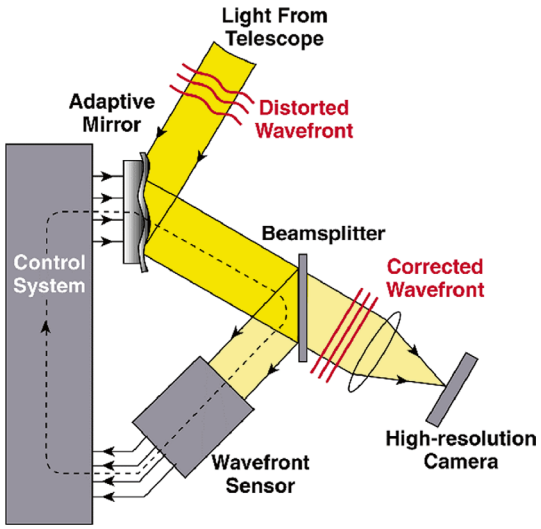


Fig. 2.28 Simplified scheme of adaptive optics (https://www.bing.com/images/search?q=adaptive+optics&qpvt=adaptive+optics&tsc=ImageHoverTitle&form=IGRE&first=1)

varies with the tilting angle of the mirror, when the telescope is directed to another object. This changes the imaging properties of the mirror and should be therefore avoided. The technique of active optics provides the solution of this problem. In order to reduce the weight of the mirror, it is made very thin (still the weight of the 8.2 m mirror on the Paranal is 23.000 kg). To its backside, many piezo-elements are attached. In Fig. 2.27 the backside of one of the primary mirrors of the very large telescope of the ESO on the Paranal illustrates the many piezo-elements, which change their length when applying a voltage across the two sides of these elements and which compensate any deformation of the mirror. The applied voltage is controlled by a computer aided feedback system which optimizes the voltage in such a way that the imaging of the primary mirror is always optimum [11].

2.6.6 Adaptive Optics

The deterioration of the star images due to air turbulences in the atmosphere can be eliminated by an ingenious technique, called "*Adaptive Optics*" [12]. Its basic principle can be explained as follows (Fig. 2.28).

The incident radiation with a distorted wavefront falls onto a thin deformable mirror. The deformation can be controlled by piezo-elements on the backside. The reflected beam passes a beam splitter. Part of it enters a wavefront sensor (Fig. 2.29) which delivers an output signal to a control system until the wavefront is an undistorted plane. The wavefront sensor consists of many micro-lenses which image the different partial beams of the incident light onto a fast CCD camera. In case of a distorted wavefront, the images of the micro-lenses are unevenly spaced. This generates a correction signal, sent to the control system until the images are evenly spaced. In this case, the star image on the high resolution camera is a diffraction limited spot.

Since the air turbulences change the refractive index within fractions of a second, the correction system must be fast enough in order to effectively correct for the distortions. The light received from an ordinary star is not bright enough to allow a fast feedback system for the correction. Therefore an artificial star is created which is by several orders of magnitude brighter than normal stars. This artificial star is created by sending a collimated laser beam with a wavelength $\lambda = 589$ nm through the telescope into the sky. In the altitude range of about 60 km a sodium layer exists, which absorbs the laser beam on the transition $3^2S \rightarrow 3^2P$ of the Na-atom. The yellow sodium line emitted by the excited Na-atoms, along the collimated laser beam, is observed through the telescope as a brilliant nearly point-like light source which is now used for driving the correction system of the passive optics (Figs. 2.30).

2.6.7 Large Telescopes

During the last years, extremely large earth-bound telescopes have been constructed with mirror diameters $D > 10$ m. Since the light gathering power is proportional to D^2 such telescopes are able to detect galaxies at far distances up to 10 billion light years, although the received radiation power is extremely weak.

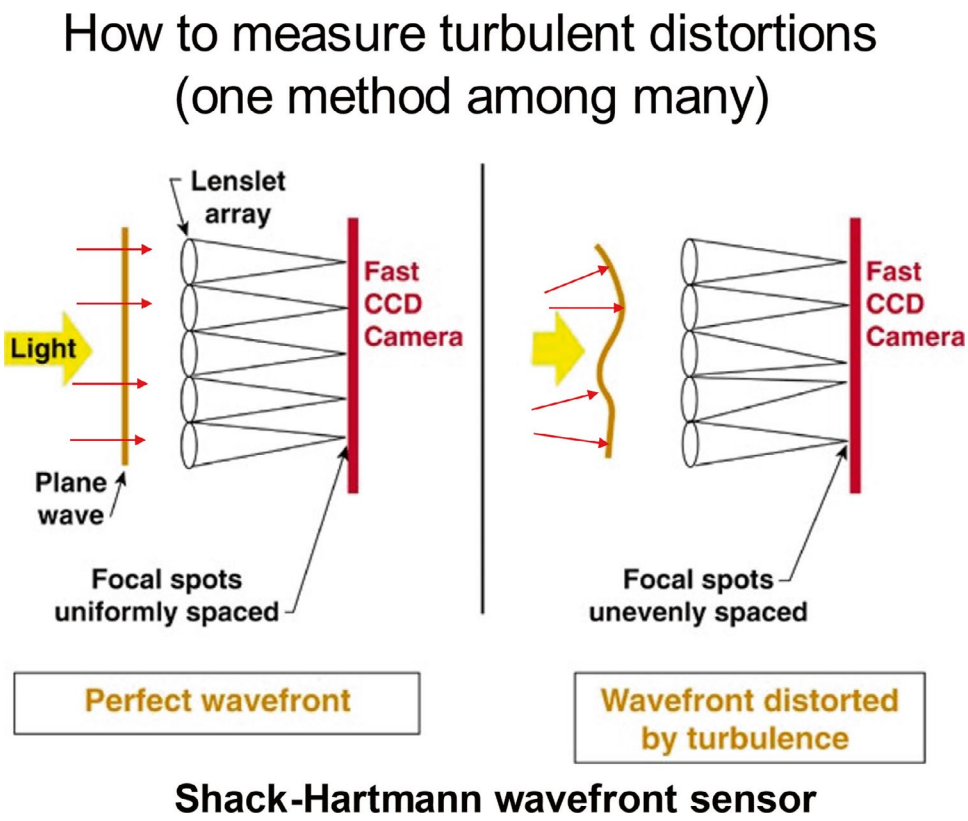


Fig. 2.29 Wavefront sensor (https://www.bing.com/images/search?q=shack-hartmann+wavefront+sensor)

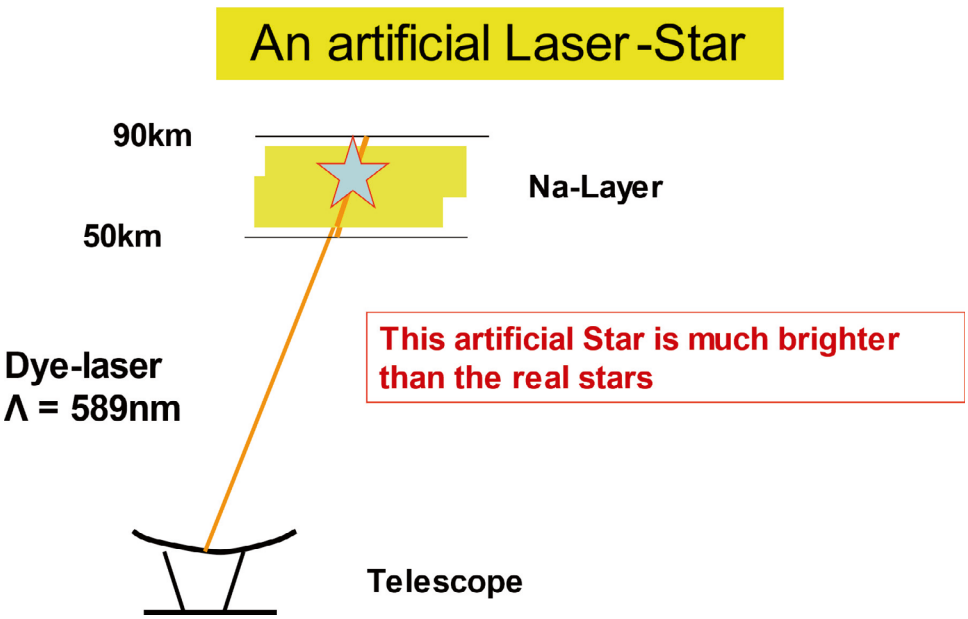


Fig. 2.30 Creation of an artificial bright star (Wikipedia)

Example

A mirror with $D = 10$ m receives (10^6 cm^2/25 cm^2) = 40.000 times more radiation power than Galilei's telescope with $D = 5$ cm. ◀

The surface quality of such a mirror with D = 10 m must be extremely high in order to minimize scattering losses. The deviation of the surface from the reference surface (e.g. a paraboloid) is for the large telescope on the Paranal less than 8.5 nm, i.e. the surface quality demands a relative roughness $\Delta\lambda/\lambda < 0.015 \rightarrow \Delta\lambda = 0.015\ \lambda$. (For $\lambda = 500$ nm this corresponds to 7.5 nm!!). In combination with active and adaptive optics the perturbing influence of air turbulence and the deformation of the mirror when changing the

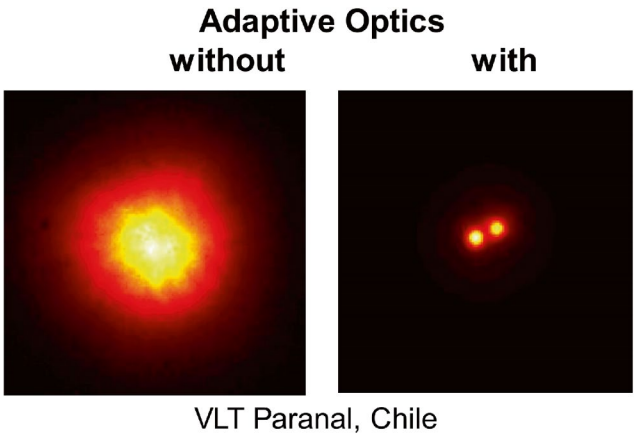


Fig. 2.31 Exposure of a binary star with and without adaptive optics (ESO) [Wikipedia]

direction of the telescope are essentially avoided. Therefore the diffraction limited angular resolution is nearly realized. This can be seen in Fig. 2.31, where two exposures of the same binary star system is shown in order to see the difference with and without adaptive optics.

During the next years a new telescope with an extremely large mirror (D = 40 m) will be installed on the Cerro Amazonas in the Chilenic Andes close to the Paranal). Such large mirrors cannot be made of a single piece. Therefore segmented mirrors are used, which consist of many hexagonal segments (Fig. 2.32). For the 40 m mirror, 800 segments form together on the parabolic surface [13]. The segments are aligned by an interferometric laser technique, which results in a nearly perfect total parabolic surface. It deviates from the reference parabolic surface by less than $\lambda/10$. This means for $\lambda = 500$ nm any deviation must be less than 50 nm! This demands a perfect precision at the fabrication and mounting of the segments. For illustration in Fig. 2.33, the segmented mirror of the 11 m telescope in South Africa is shown.

In Table 2.2 the largest worldwide realized telescopes are compiled, while in Table 2.3 the largest telescopes planned in the near future or already under construction are listed (Fig. 2.34).

2.6.8 Telescope Mounts

In order to keep the direction of a telescope for a longer time always towards a selected point on the sky, the telescope mount must be sufficiently stable and must be able to exactly compensate for the rotation of the earth. The most simple way to fulfill these requirements is the **parallactic mount** (also named **equatorial mount**) shown in Fig. 2.35. The telescope is turned around an axis (hour axis) parallel to the earth's axis by an electronic control system. The rotation rate is 15° per siderial hour, opposite to the earth's rotation. The telescope can be turned around a second axis (declination axis) perpendicular to the hour axis in order to direct it towards the desired object.

The weight of the telescope is compensated by a corresponding counterweight, which balances the torque of the telescope around its axes (Fig. 2.36). For the proper design of this counterweight, the total torque at each telescope direction is nearly zero and allows the easy change to any wanted direction. If the center of gravity of the telescope is at the intersection of the two axes, no counterweight is needed, because the total torque is always zero.

Another possibility is the **altazimuth mount** where one axis is vertical and the other horizontal (Fig. 2.36b). Its advantage is the simple design. Rotation around the vertical axis changes the azimuth of the pointing direction while

Fig. 2.32 Composition of a segmented parabolic mirror by hexagonal segments (https://www.bing.com/images/search?q=large+segmented+telescope+mirrors)

Fig. 2.33 Segmented primary mirror of the South African Large Telescope [12]

rotation around the horizontal axis allows to direct the instrument to stars with different elevations.

Its disadvantage is the fact that the telescope has to be turned simultaneously around both axes in order to follow the selected star, while the earth rotates around the earth's axis.

Nevertheless, modern large telescopes use this *altazimuth mount* because the cost savings of the much simpler mount are essential. Since the pointing direction of the telescope is nowadays always controlled by a computer program, the simultaneous rotation around two axis really does not represent any disadvantage.

2.6.9 Radio Telescopes

We receive from the universe not only the visible light from stars, but a broad spectrum of electromagnetic radiation, ranging from the radio-frequency region to γ-radiation. From this spectrum the earth's atmosphere transmits only some narrow windows (see Fig. 2.1). The radio-frequency range belongs to one of these transmission windows.

Beginning with the end of the 1930 years, but on a larger scale only after World War II the construction of radio telescopes started with different constructive modifications. The version of radio telescope with directional antennas can be directed onto any point on the sky within the accessible angular range. For many years the radio telescope in Effelsberg. Germany (Fig. 2.37) was worldwide the largest directional radio telescope [14] with a paraboloid diameter of 100 m (Table 2.4) [8]. In 2000 the Robert C.B Byrd radio telescope with a paraboloid diameter of 110 m was built in Greenbank, West Virginia USA [15].

The radio-radiation from the selected region of the sky is collected by the paraboloid antenna and focused onto the detector, which is a sensitive semiconductor element cooled down to 4.2 K. The exit signal is mixed with the output of a local oscillator (heterodyne principle) and the difference signal is sent through a frequency filter which suppresses the background noise. Tuning the frequency of the local oscillator any wanted radio frequency is transmitted through the fixed frequency filter [9].

Where does the radio frequencies come from? The most important radio signals are emitted by hydrogen atoms at $\lambda = 21\,\text{cm}\,(\nu = 1.43\,\text{GHz})$ on the transition between the two hyperfine levels $F = 1 \rightarrow F = 0$ in the 1 s electronic ground state. Many other frequencies are emitted by rotational transitions of interstellar molecules.

The 64 m radio telescope at Parkes Observatory (Fig. 2.38) was built in 1969. It has been used, (besides its main purpose to measure radio-frequency radiation from the sky) to receive televised videos from the Apollo mission during the first landing of man on the moon.

There are several modifications of Radio Telescopes: Besides the directional antenna telescopes, a large variety of stationary telescope arrays exist. The largest one was the Arecibo Radio telescope in Puerto Rico with a parabolic receiver diameter of 300 m, which was built into a valley (Fig. 2.39). It rotates with the earth and can therefore receive signals from all radio sources which are acceptable within the angular declination range Δδ of the telescope. The detector was fixed on steel cables above the parabolic dish. It could be shifted within small ranges, which enlarges the receivable declination range Δδ slightly (Arecibo Observatory).

Unfortunately, the detector support was destroyed in 2020 during a thunderstorm. It fell down onto the parabolic dish and damaged it so severely, that a repair was not reasonable. There is now an even larger radio telescope with a stationary parabolic dish with 500 m diameter in China (Fig. 2.40) which was finished in 2016 and represents the largest radio telescope in the world [16].

The angular resolution can be greatly enhanced if several telescopes with a diameter d and a maximum distance D between the first and last telescope can be synchronized (Fig. 2.41), which means that the signals of the different telescopes are recorded at the same time, synchronized by atomic clocks (Fig. 2.43). The combined signals represent an intensity correlation. Now no longer the diameter d of a single telescope limit the angular resolution but only the much larger distance D. This can be seen as follows (Fig. 2.42).

The radio-radiation with the wavelength λ from a far distant source arrives at the telescopes as a plane wave. The path difference from a source Q_1 to the telescopes T_1 and T_2 with the mutual distance L is

$$\Delta s_1 = Q_1T_1 - Q_1T_2 = L \cdot \sin\alpha_1 \qquad (2.7)$$

Table. 2.2 The worldwide largest telescopes (Wikipedia)

Name	Effective aperture	Mirror type	Nationality/sponsors	Site	First light
Large Binocular Telescope (LBT)	11.9 m (469 in) (combined)	Multiple Two 8.4 m (331 in) mirrors	USA, Italy, Germany	Mount Graham International Observatory, Arizona, USA	2004
Gran Telescopio Canarias (GTC)	10.4 m (409 in)	Segmented 36 hexagonal segments	Spain, Mexico, USA	Roque de los Muchachos Obs., Canary Islands, Spain	2006
Hobby-Eberly-Telescope (HET)	10 m (394 in) (effective)	Segmented 91 × 1 m (39 in) Hexagonal segments forming a 11 m × 9.8 m	USA, Germany	McDonald Observatory, Texas, USA	1997 Aperture increased
Keck 1	10 m (394 in)	Segmented 36 hexagonal segments	USA	Mauna Kea Observatories, Hawaii, USA	1993
Keck 2	10 m (394 in)	Segmented 36 hexagonal segments	USA	Mauna Kea Observatories, Hawaii, USA	1996
Southern African Large Telescope (SALT)	9.2 m (362 in) (effective)	Segmented 91 × 1 m (39 in) hexagonal segments forming a 11 m × 9.8 m mirror	South Africa, USA, UK, Germany, Poland, New Zealand	South African Astronomical Obs., Northern Cape, South Africa	2005
Subaru (JNLT)	8.2 m (323 in)	Single	Japan	Mauna Kea Observatories, Hawaii, USA	1999
VLT UT1—ANTU	8.2 m (323 in)	Single	ESO Countries, Chile	Paranal Observatory, Antofagasta Region, Chile	1998
VLT UT2—Kueyen	8.2 m (323 in)	Single	ESO Countries, Chile	Paranal Observatory, Antofagasta Region, Chile	1999
VLT UT3 – Melipal	8.2 m (323 in)	Single	ESO Countries, Chile	Paranal Observatory, Antofagasta Region, Chile	2000
VLT UT4 – Yepun	8.2 m (323 in)	Single	ESO Countries, Chile	Paranal Observatory, Antofagasta Region, Chile	2001

If the two signals are mixed with the local oscillator, the two resulting output signals have a phase difference

$$\Delta\varphi_1 = (2\pi/\lambda) \cdot \Delta s_1 \tag{2.8}$$

For the radiation from a neighboring source Q_2 the analogue relations are

$\Delta s_2 = L \cdot \sin\alpha_2$ and $\Delta\varphi_2 = (2\pi/\lambda)\Delta s_2$.

The difference of the two phase shifts is

$$\delta\varphi = \Delta\varphi_1 - \Delta\varphi_2 = (2\pi/\lambda)(\Delta s_1 - \Delta s_2) \tag{2.9}$$

This phase shift difference should be smaller than π, because for $\delta\varphi = \pi$ the amplitude maximum of the signal from Q_1 superimposes the minimum for Q_2. This is the Rayleigh-condition that the two signals can be just distinguished (see Vol. 2, Sect. 11.5.3 and Fig. 2.25).

With $\Delta\alpha = \alpha_1 - \alpha_2 \ll 1$ we obtain

$$\begin{aligned} &\Delta\varphi_1 - \Delta\varphi_2 = (2\pi/\lambda) \cdot L \cdot (\sin\alpha_1 - \sin\alpha_2) < \pi \\ &\sin\alpha_1 - \sin\alpha_2 = \sin\alpha_1 - \sin(\alpha_1 - \Delta\alpha) \\ &= \sin\alpha_1 - (\sin\alpha_1 \cdot \cos\Delta\alpha - \cos\alpha_1 \cdot \sin\Delta\alpha) \\ &\approx \cos\alpha_1 \cdot \Delta\alpha \end{aligned} \tag{2.10}$$

The minimum still resolvable angle α is then

$$\Delta\alpha \geq \frac{\delta\varphi \cdot \lambda}{2\pi \cdot L \cdot \cos\alpha_1} \geq \frac{\lambda}{2L}. \tag{2.11}$$

Table. 2.3 Large telescopes planned or under construction

Name	Aperture diameter (m)	Collecting area (m^2)	First light
Extremely Large Telescope (ELT)	39.3	978	2025
Thirty Meter Telescope (TMT)	30	655	2027
Giant Magellan Telescope (GMT)	24.5	368	2029

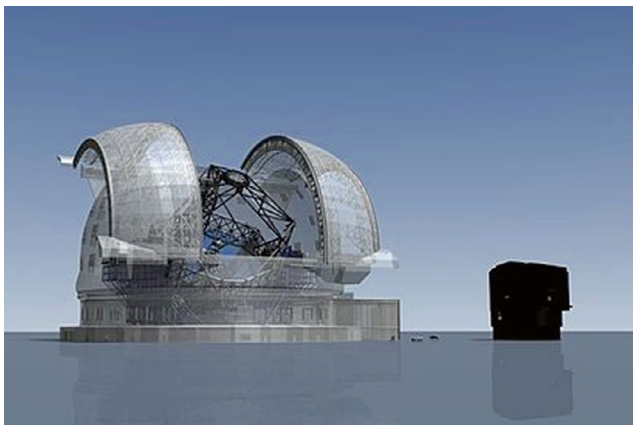


Fig. 2.34 Size of the extremely large 40 m Telescope under construction compared with the existing VLT (8.2 m) on the Paranal (ESO) the small structure on the right site is the already existing verify large telescope

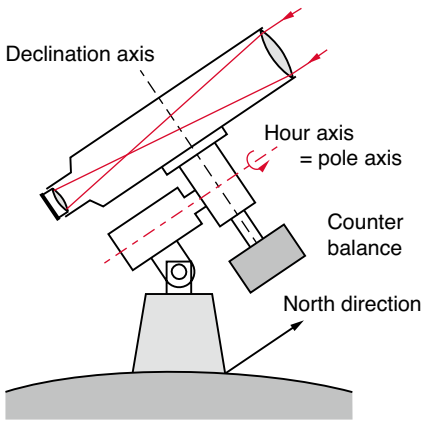


Fig. 2.35 Parallactic mount of a telescope

The angular resolution increases with the length L of the baseline between two telescopes.

Such long baseline radio interferometry (more accurately it should be named **radio correlation**, because it is no interferometry in the usual meaning) can therefore increase the angular resolution by many orders of magnitude [18].

Example

One radio telescope is located in Germany (e.g. in Effelsberg) the second telescope in the USA. The baseline is then about 6000 km. This increases the angular resolution for $\lambda = 21$ cm and $L = 6 \times 10^6$ m to

$$\Delta\alpha_{\min} \approx \frac{\lambda}{2L} \leq 2 \cdot 10^{-8} \approx -0.''0035$$

which exceeds the angular resolution of large optical telescopes. ◀

For such large distances between the two telescopes, one cannot use the same local oscillator. One uses instead two separate oscillators which are synchronized by atomic clocks. The intensities of the mixed signals recorded by the two telescopes are used to form the signal as the intensity correlation [17, 18]

$$S = \int_0^\infty \frac{I_1(t) \cdot I_2(t + \Delta t)\Delta t}{I_1(t) \cdot I_2(t)} \tag{2.12}$$

The whole procedure is illustrated in Fig. 2.43.

2.6.10 Optical Interferometry in Astronomy

The combination of two or more optical and near-infrared telescopes to form an interferometer has been recently realized. Because of the much shorter wavelength λ than in Radio-Astronomy ($\lambda_{opt} = 10^{-5}\ \lambda_{radio}$) the demands with regard to accuracy and stability of the distance D between the two telescopes are much more severe. The path difference between the interfering light beams should not fluctuate by more than $\lambda/10$, which means for $\lambda = 500$ nm by no more than 50 nm! For a telescope distance $D = 100$ m this implies a maximum relative distance change of less than $\Delta D/D < 5 \times 10^{-10}$!!!

The first stellar interferometer has been already realized in 1928 by A. A. Michelson and co-workers (see Sect. 4.1.1) who measured the radius of giant stars. Its

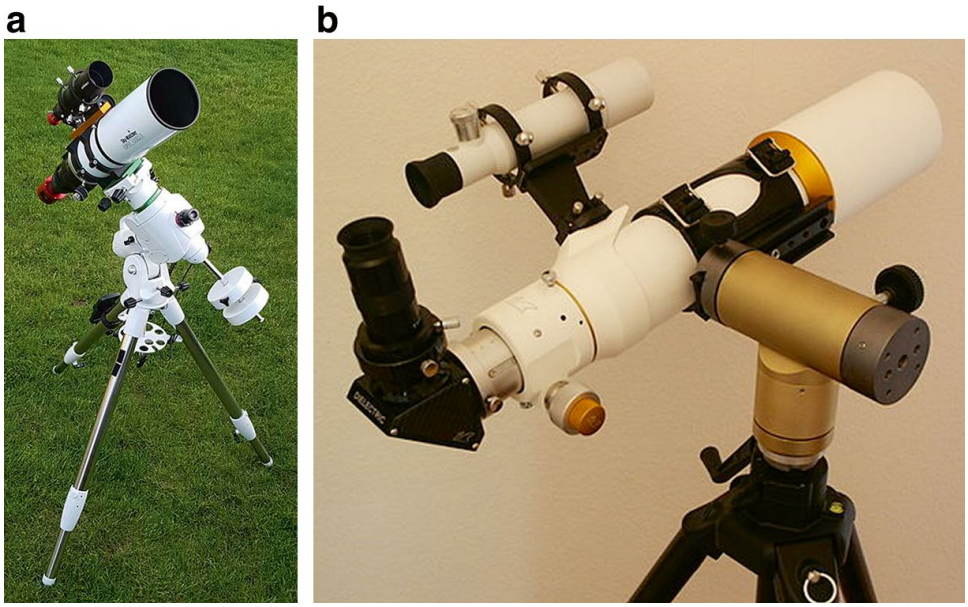


Fig. 2.36 **a** Small telescope with equatorial mount and counterweight for amateur sky observation. **b** Telescope with azimutal mount

Fig. 2.37 Radio telescope in Effelsberg, Germany

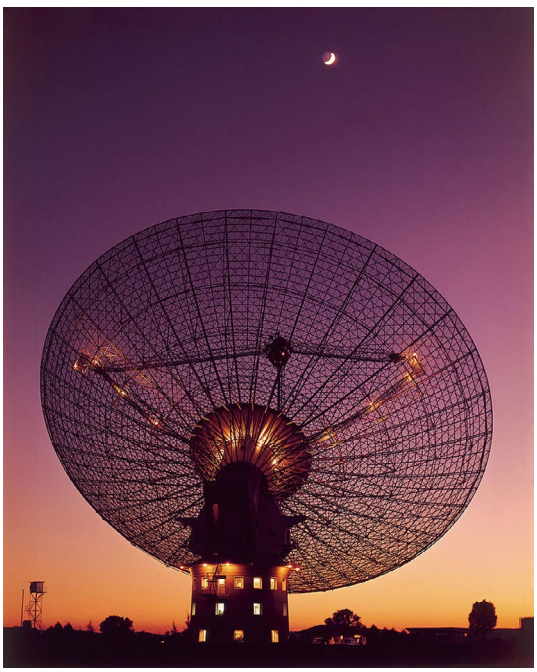


Fig. 2.38 The 64 m radio telescope at Parkes Observatory (By CSIRO, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=35481714)

principle is illustrated in Fig. 4.2. In these early times the maximum distance was limited to $D = 6\,\text{m}$, because for larger distances the air turbulence blurred the interference fringes and completely smeared out the interference structure during the necessary exposure time.

The application of adaptive optics (see Sect. 2.6.6) could significantly eliminate the influence of air turbulence and nowadays it is possible to combine large telescopes (for instance the ESO- telescopes on the Paranal) with distances up to 300 m to a large optical interferometer. The

Table. 2.4 Technical data of the radio telescope Effelsberg in Germany

Construction	1968–1971
Mirror diameter	100 m
Surface of reflector	7850 m^2
Surface flatness	< 0.5 mm
Mass	3200 tons
Usable wavelength range	$\lambda = 43.5$–900 mm
Angular resolution (at $\lambda = 21$ cm)	$10''$

Fig. 2.39 Arecibo-radio telescope in Puerto Rico (Arecibo Observatory) which was unfortunate destroyed by a storm, where the detector fall down onto the parabolic disc

Fig. 2.40 Large radio telescope with a 500 m diameter spherical disc in Guizhou, China (Wikipedia)

Fig. 2.41 Very large interferometric array of 27 parabolic radio telescopes in Sorroco, New Mexico (Wikipedia)

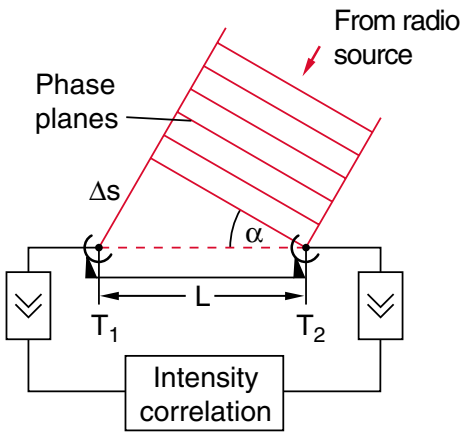


Fig. 2.42 Schematic illustration of the increased angular resolution of two synchronized radio telescopes with baseline L

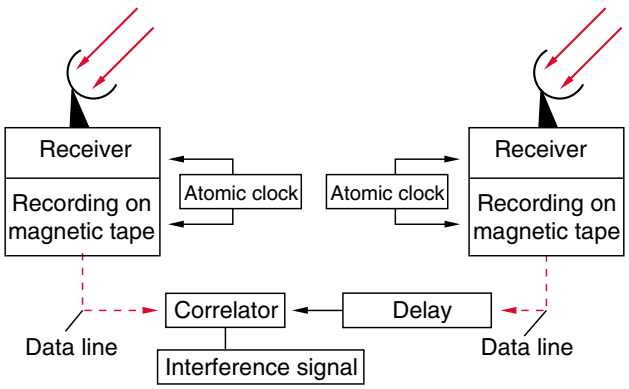


Fig. 2.43 Schematic arrangement of the very long baseline interferometry

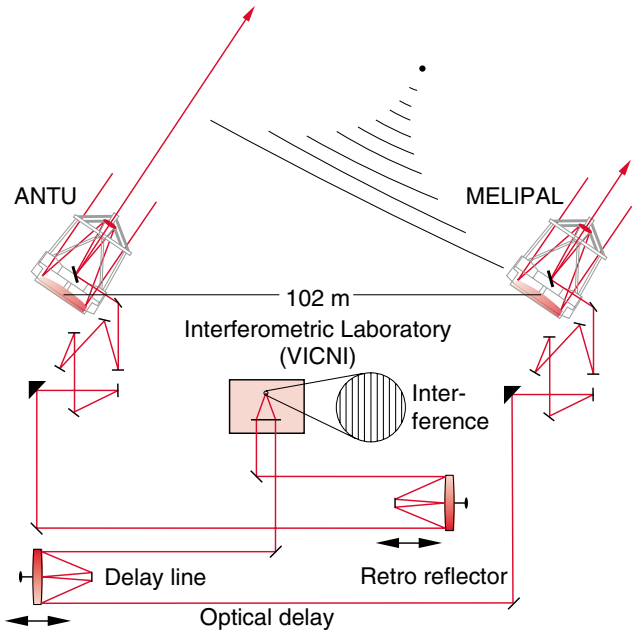


Fig. 2.44 Principal setup of the optical interferometer with two interconnected large telescopes of the ESO on the Paranal in Chile

experimental efforts and the sophisticated technique are impressive [19].

The optical beam path is schematically shown in Fig. 2.44. The beam line with many mirrors proceeds in a tunnel between the two telescopes which are named after Indian deities. The path difference can be continuously varied by retroreflecting prisms mounted on a carriage which slides vibration-free on polished rails (Fig. 2.45).

Fig. 2.45 Part of the optical delay line realized by a retroreflector mounted on a carriage smoothly rolling on polished rails (ESO/H. Zodet)

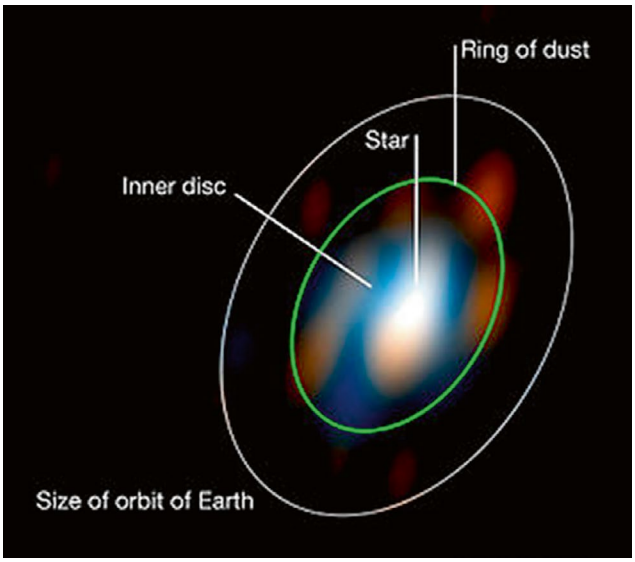


Fig. 2.46 The first image of an airy disc around a star during its formation was taken by the interferometric combination of two large telescopes at the Paranal [15]

The angular resolution of such interferometrically combined telescopes corresponds to that of a single telescope with a mirror diameter of D = 300 m.

In Fig. 2.46 a picture of a star during its formation is shown, taken with the large telescope interferometer of the ESO on the Paranal. The ring of dust and the inner disc around the young star which has not yet condensed into the star can be clearly seen.

2.6.11 X-ray Telescopes

Many sources in the universe (also our sun) emit X-rays. They are absorbed in our atmosphere. Therefore X-ray detectors in Satellites outside our atmosphere have to be used in order to investigate these X-rays, their intensity and spectral distribution. One of the most successful X-ray Satellites was the ROSAT [20], built in the Max-Planck Institute in Garching under the directorship of Prof. Trümper, which was launched in 1990 and was active from 1990 to 1999. During this active period, it discovered 125.000 new X-ray sources [21].

The principle of an X-ray telescope is illustrated in Fig. 2.47. The incident X-rays hit rotationally symmetric parabolic mirror walls under a small grazing incidence angle $\alpha < 3°$ (because the reflectivity of X-rays increases with decreasing incidence angle). Because of the small wavelength ($\lambda = 12.5$–0.06 nm corresponding to the X-ray quantum energy of 100 eV to 20 keV). the surface of the

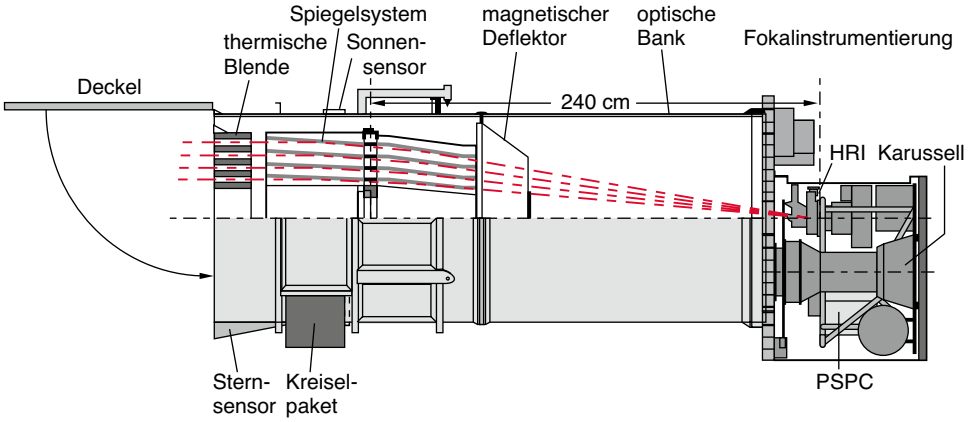


Fig. 2.47 The X-ray telescope ROSAT (from B. Aschenbach, H.-M. Hahn, J. Trümper: Der unsichtbare Himmel. Birkhäuser Basel)

reflecting mirror must be extremely smooth polished. Even a small roughness would result in scattering of the X-radiation, which leads to a deterioration of the imaging quality.

The detectors are two proportional counters with high spatial resolution and micro-channel plates with wedge shaped anode stripes, which work even under intense incident radiation without problems. The angular resolution of ROSAT is 4″ and the accepted field of view is a circle with 2° diameter.

New X-ray Satellites are *Chandra* (named after the Indian Astrophysicist *Subrahmany Chandrasekhar*) and *Newton*, which had several improvements compared with ROSAT. For instance, Chandra had a better angular resolution of 0.5″ but a smaller field of view (30′ × 30′). Its main advantage is the higher spectral resolution, which was achieved with pn-CCD-detectors. These detectors are integrated arrangements of photon detectors with a good space- and spectral- resolution (Fig. 2.48).

Meanwhile, more than 50 X-ray Satellites have been launched where more than 10 are still in operation. A special task of some of them is the detailed observation of the X-radiation from our sun. Therefore a satellite **SOHO** (**s**olar **h**elioscopic **o**bservatory) was launched which is kept in the Lagrange point L1 between earth and sun and which orbits around the sun synchronously with the earth. It observes all dynamic processes, such as sun spots, flares, eruptions, protuberances and sun spots over a broad spectral range including the X-ray region.

The NASA satellite IRIS (**I**nterface **R**egion **I**maging **S**pectrograph Fig. 2.49) observed in detail the sun limb, in particular the corona of the sun in the X-ray region up to the UV range. One of its major instruments is a high rate imaging ultraviolet spectrometer. The detection of X-rays allows the study of solar eruptions in the very beginning, before they show up in the visible range.

Fig. 2.48 Mounting of the CHANDRA X-ray satellite (NASA)

The multi-wavelength satellite *Astrosat* was launched in 2015. Due to the broad spectral range of its spectrometer, it is able to observe astronomical objects simultaneously in the whole spectral region from the visible to the X-ray range.

2.6.12 Detectors for Cosmic Gamma Radiation

The electromagnetic radiation in the energy range above 100 keV is called **Gamma Radiation**. There are many possible sources for the cosmic gamma radiation. One of these sources is radioactive isotopes which have been formed in

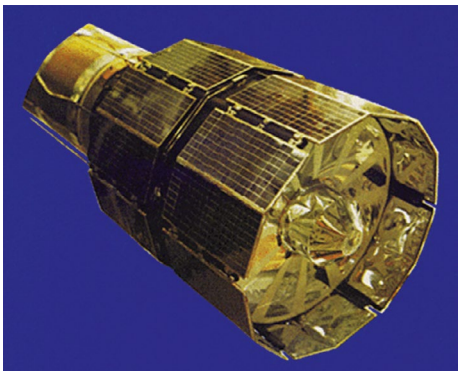


Fig. 2.49 X-ray NASA solar observation satellite IRIS (Wikipedia)

Fig. 2.50 HESS telescopes in Namibia (HESS, picture gallery)

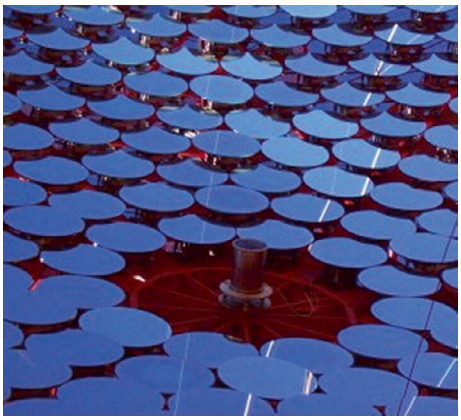


Fig. 2.51 The mirror surface of a HESS-telescope is composed of many small reflecting plates (HESS picture gallery)

the interior of stars during the nuclear fusion processes (see Chap. 4) and are ejected during supernova explosions of stars.

Among these radioactive isotopes are also β^+-emitters. By collisions of these positrons with the electron shell of atoms in the earth's atmosphere annihilation processes $\beta^+ + \beta^- = 2\gamma$ occur where 2 gamma-quanta with energies of 512 keV each are produced.

Another source of γ-radiation is neutron stars which emit γ-quanta during the conversion of protons into neutrons and also stars, that spiral into black holes (see Chap. 6). A particular intense source of γ-radiation is supernova explosions.

The γ-quanta reaching the earth can be detected either directly by detectors in satellites outside the earth-atmosphere, or indirectly by earth-bound detection systems, which mainly measure the secondary radiation induced by the γ-quanta in the atmosphere.

Besides the X-ray satellites the first γ-radiation satellite was launched in 1990 in Russia, which was followed in 2008 by the essentially improved *Gamma Radiation Laboratorium* FGRST (Fermi Gamma Ray Space Telescope, NASA) which could measure the energy of γ -quanta in the range between 10 meV and some GeV.

Among the earth-bound γ-radiation detectors, the two stations, the HESS Observatory (High Energy Stereoscopic System) in Namibia (Figs. 2.50 and 2.51 and [23]) and the MAGIC Observatory (Major Atmospheric Gamma Imaging Cerenkov Telescope) on the Canarian Island *La Palma* (Fig. 2.52) are successfully operating since 2004.

The HESS observatory consists of an array of imaging Cherenkov telescopes (Fig. 4.50) which allow the stereoscopic observation of air showers produced in the earth's atmosphere with a maximum production rate at about 10 km altitude by incident γ-rays in the 10 GeV to 10 TeV energy range. Since the shower particles fly with nearly the speed of light, they generate Cerenkov radiation, because the light travels in air with refractive index n > 1 with the velocity $c = c_0/n$ which is smaller than the vacuum speed c_0 This Cerenkov radiation is collected by the HESS-telescopes. The parabolic mirror is composed of 380 circular mirrors with 60 cm diameter each (Fig. 2.51).

As one recent example of astronomical results is the measurement of the spatial extension of the γ-emitting area in the radio-galaxy Centaurus A (NGC 5128) which emits γ -quanta in the TeV-range. Also, the question of how these high energy γ-rays are produced, could be solved.

The diameter of the reflecting parabolic mirror of MAGIC is 17 m, and the detected energy range of the γ -quanta is 25 GeV to 30 TeV. The aim of the research is the investigation of γ-radiation from remnants of supernova explosions or from the formation of supermassive black holes in the center of active galaxies [22].

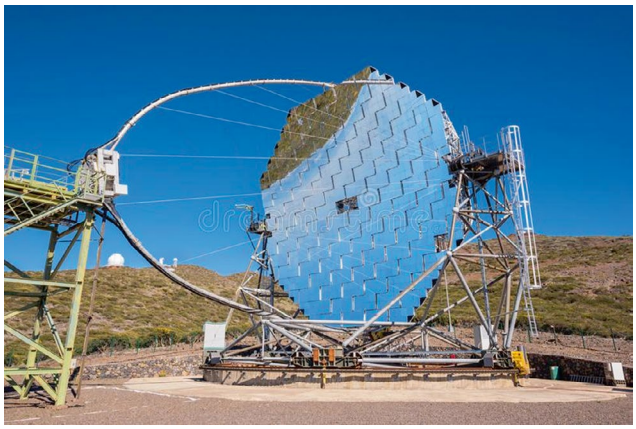


Fig. 2.52 MAGIC Telescope on Las Palmas (https://en.wikipedia.org/wiki/MAGIC_telescope

In Fig. 2.53 the signals from a supernova explosion are shown as detected by MAGIC. It demonstrates that the time profiles of the detected signals are different for the different spectral ranges.

2.6.13 High Energy Cosmic Radiation

The Cosmic radiation was first discovered in 1912 by Victor Hess (Fig. 2.54) who made balloon ascensions up to 5.3 km altitude and used three electrometers to detect ionizing particles with very high energy (Nobel Prize 1936).

The primary Cosmic radiation consists of very high energy (up to 10^{20} eV) particles (protons, electrons and He-nuclei), reaching the earth from Cosmic sources. When they enter the earth's atmosphere, new particles are generated (secondary radiation, consisting of pions, muons, γ -quanta) by collisions of the high energy particles with atoms and molecules of the atmosphere (Fig. 2.55) One proton with the energy $E = 10^{15}$ eV can produce more than 10^6 secondary particles. In collisions of protons with atomic nuclei N the following reactions can occur:

$$\mathrm{p} + \mathrm{N} \rightarrow n_1 \cdot \mathrm{p} + n_2 \cdot \mathrm{n} + n_3 \cdot \pi + n_4 \cdot \mu \qquad (2.13)$$

where the relative shares of the different secondary products depend on the energy and species of the primary particles. The unstable secondary particles decay, according to

$$\pi_0 \rightarrow \gamma + \gamma;\ \pi^- \rightarrow \mu^- + \overline{\nu}_\mu;\ \pi^+ \rightarrow \mu^+ + \nu_\mu \qquad (2.14a)$$

$$\mu^+ \rightarrow \mathrm{e}^+ + \nu_\mathrm{e} + \nu_\mu;\ \mu^- \rightarrow \mathrm{e}^- + \nu_\mathrm{e} + \nu_\mu \qquad (2.14b)$$

The rate of cosmic ray particles decreases strongly with increasing energy (Fig. 2.56).

For the detection of the cosmic radiation, several techniques are used. Either the secondary particles are directly detected or the Cerenkov radiation produced by these fast particles is measured (see Sect. 2.6.12).

Meanwhile, several large detector fields have been built for the measurements of such particle showers. One of them is the Pierre Auger Observatory which covers an area of 3.000 km^2 in the Pampas in Argentina. Since the high energy particles reach nearly the velocity c of light, they produce Cerenkov radiation and can be detected by Cerenkov detectors (see Vol. 4, Sect. 4.3.5). The fluorescence which is produced in collisions of the particles

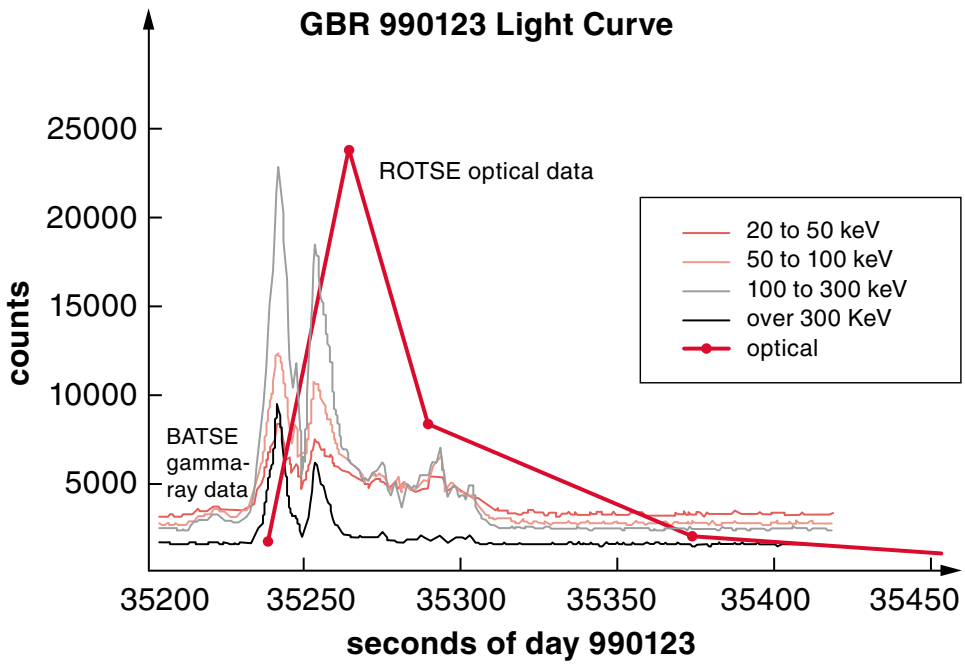


Fig. 2.53 Comparison of the different signals from a supernova explosion measured at different wavelengths in the visible and gamma spectral range (http://magic.mppmu.mpg.de/introduction/index.html)

Fig. 2.54 Victor Hess (1883–1964) (Public Domain, https://commons.wikimedia.org/w/index.php?curid=258401)

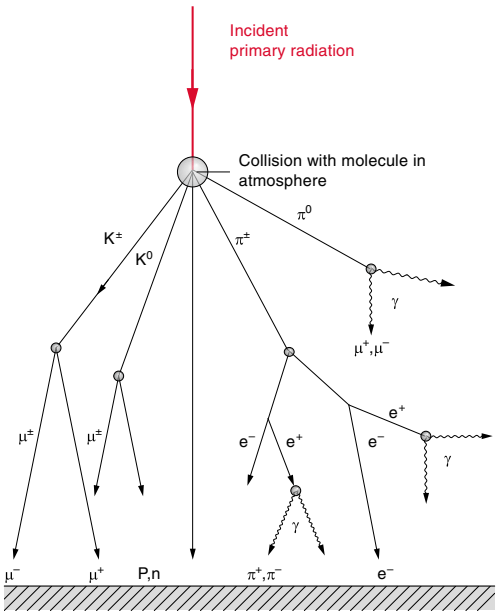


Fig. 2.55 Cascade shower of different secondary particles in the earth's atmosphere, produced by the incident cosmic rays

with molecules in the atmosphere can be detected by photomultipliers.

In 2002 the satellite *Integral* (International Gamma-ray Astrophysics Laboratory) was launched, which carries 4 telescopes with different spatial and spectral resolutions [24]. There are three possible sources of the cosmic radiation:

- Emission from the sun
- Other sources in our galaxy
- Extragalactic radiation

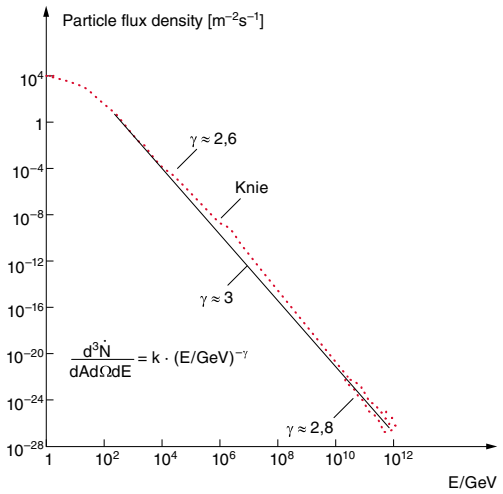


Fig. 2.56 Energy spectrum of cosmic rays. It is proportional to $E^{-\gamma}$ where γ varies for the different energy ranges between 2.6 and 3

Hess could exclude the sun as a possible source for the high energy particles, when he made a balloon ascension during a total solar eclipse, where he still found the cosmic rays. The sun does emit particles during solar eruptions like flares and protuberances. But these particles do not have the extremely high energy, detected in the cosmic rays. One possibility for the generation of high energy particles is the acceleration of charged particles in the center of active galaxies, where black holes attract matter and accelerate it to very high energy. Another possible source is supernova explosions of massive stars [25].

2.7 Detectors for Gravitational Waves

Analogous to the emission of electromagnetic waves by accelerated charges accelerated masses emit gravitational waves [26]. They cause a distortion of space and material bodies which experience gravitational forces if they are hit by gravitational waves. Possible sources of gravitational waves are supernova explosions or two massive bodies such as neutron stars or black holes which circle around their common center of mass and finally merge. Einstein postulated already in his general relativity theory the existence of gravitational waves, but he believed, that they could never be detected because of their weak interaction with matter.

Since the gravitational force is very weak (compared to the other forces, such as the electromagnetic force) the distortion of bodies on earth by gravitational waves is extremely small and very sensitive detectors have to be developed in order to measure these waves. First attempts had been tried by Weber and co-workers with a massive aluminum cylinder which was expected to show resonant vibrations with a very small amplitude under the influence of gravitational waves. It turned out, however, that these

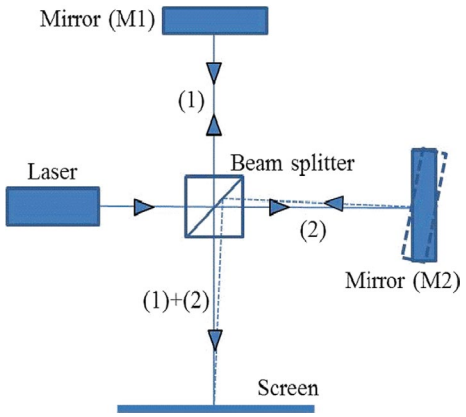


Fig. 2.57 Schematic drawing of a Michelson Interferometer

amplitudes were even smaller than the noise vibrations although very careful shielding of all possible perturbations had been assured.

Some competing groups used a laser interferometer in the form of a Michelson interferometer as a detector (Fig. 2.57). The incoming laser beam is split by the beam splitter into two arms which are perpendicular to each other. After reflection at the mirrors M1 and M2 the two partial beams are recombined and superimposed on the detector screen. The measured intensity depends on the relative phase of the two partial beams after recombination they have a phase difference which results in a structure of dark and bright stripes in the observation plane, if one of the mirrors has been slightly inclined.

An improved version of a Michelson interferometer is shown in Fig. 2.58. Here the length of the two side arms is greatly enlarged by optical resonators. This increases the sensitivity considerably.

The first successful detection of gravitational waves was possible with a sophisticated laser interferometer and was announced by the LIGO cooperation in 2014 [27].

For the detection of gravitational waves with large wavelengths an extraordinary detector is planned (Fig. 2.59) which consists of 3 Satellites located on the corners of an equilateral triangle with side lengths $L = 10^9$ m. The center of the three satellites moves on the earth's orbit around the sun, 20° behind the earth. Each satellite carries a laser source, a detector and two mirrors. The mirrors on each two satellites act as interferometers. This implies that the distance between the two mirrors has to be kept constant within a fraction of the wavelength. A gravitational wave changes the two distances which form the angle $\alpha = 60°$ against each other, by a different fraction of their length, which can be detected by the change of the resonator resonance, causing a change in the laser intensity.

Summary

- The position astronomy (astrometry) measures most accurately the angular positions (ϑ, φ) of stars and their variations in time.
- The horizon coordinate system seems more natural for the observer, but has the disadvantage that it rotates with the earth and is different for different locations on earth. The equator system is independent of the earth's rotation. Its reference plane is the plane through the earth's equator which cuts the celestial sphere in the celestial

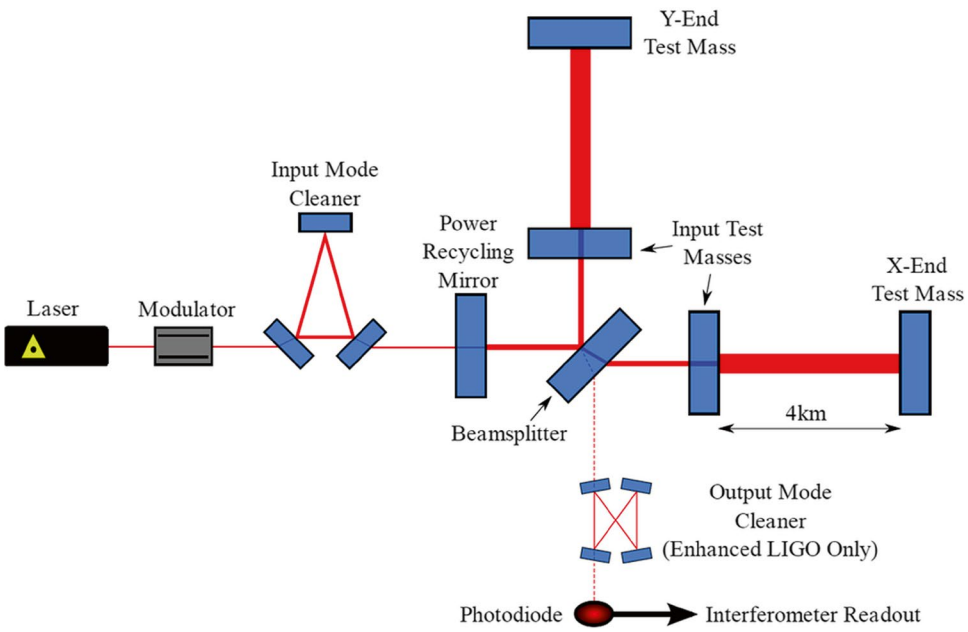


Fig. 2.58 Improved laser interferometer used by the LIGO group (by J. R. Smith for the LIGO Scientific Collaboration Wikimedia)

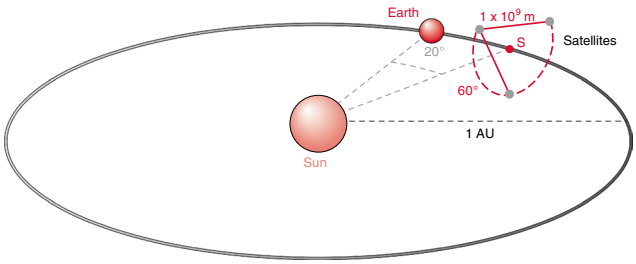


Fig. 2.59 Planned gravitational wave detector LISA

equator. The latitude through a star defines its declination δ, the meridian its rectacension α, which is measured from the vernal equinox, the intersection of ecliptic and celestial equator. The sun stands at the vernal equinox on the 21st of March.

- Due to the precession and nutation of the earth's axis, the vernal equinox shifts in 26.000 years by 2π. The direction of the earth's poles is also changing accordingly and therefore also the declination δ and the rectacension α. These changes are listed in astronomical tables.
- The orbit of a star apparent for an observer on the rotating earth is divided into the diurnal arc from the rise of a star to its setting and the night arc where it is not observable. The length of the diurnal arc depends on the declination δ of the star and the geographical position φ of the observer.
- Stars with $\delta > 90° - \varphi$ are circumpolar. They are visible during the whole year. Stars with $\delta < \varphi - 90°$ can never been seen by an observer at the latitude φ.
- For astronomical observations lens telescopes (refractors) and mirror telescopes (reflectors) are used. Their light gathering power is proportional to the square D^2 of the usable diameter D of the telescope. The angular resolution is in principle limited by the diffraction $\alpha_{\min} = 1.22 \cdot \lambda/D$. The air turbulence deteriorates the resolution to $a \approx 0.5''$–$1''$ (seeing). This can be overcome by special techniques such as the adaptive optics. For telescopes on satellites outside the atmosphere, the diffraction limit can be nearly reached.
- In order to compensate for the earth's rotation, telescopes need an auto-tracking which is controlled by a computer feedback and guarantees that the telescope is directed during the whole exposure time to a selected point on the sky.
- Besides telescopes for the visible and near infrared spectral range radio telescopes which detect microwave and radio-radiation ($\lambda = 350\,\mu\text{m}$ to 1 m) have contributed much to our knowledge about the structure and evolution of the universe. The interconnections of several telescopes with large mutual distances (long baseline interferometry) have increased the angular resolution above that of visible telescopes.
- Infrared telescopes ($\lambda = 0.8$ to 10 μm) and X-ray telescopes mounted on Satellites outside the earth's atmosphere, have greatly extended the spectral ranges and have increased the information about processes in the universe.
- The interconnection of several visible telescopes by optical paths with more than 100 m length has enabled interferometry with a very high spectral and angular resolution.
- Cosmic rays cover a wide range of energies up to 10^{21} eV. They are detected by measuring the Cerenkov radiation of the secondary particles produced by collisions of the primary high energy particles with atoms and molecules in the earth's atmosphere.

Problems

2.1 Derive the equations of the transformation from the Horizon system to the equator system.

2.2 Determine the duration of the longest and the shortest day (from sun rise to sun set) at the latitude $\varphi = 50°$. The inclination angle of the ecliptic is 23.5°.

2.3 What is the diameter of the apparent elliptic relative motion of two stars (parallax) with distances $r_1 = 3$pc and $r_2 = 6$pc?

2.4 Calculate the escape velocity of a body from the surfaces of Venus, Jupiter, earth moon and the sun.

2.5 A laserpulse (peak power $P = 1$ GW, $\lambda = 500$ nm duration $\tau = 150$ ps) is sent through a telescope ($D = 0.75$ m) to the moon.

(a) What is the diameter of the central diffraction maximum on the moon under ideal conditions?

(b) The real beam divergence due to diffraction and seeing id $\alpha = 3''$. Which fraction η of the emitted photons reaches the retroreflector (0.4×0.4 m^2) on the moon?

(c) Which fraction η_2 of the reflected laser power reaches the telescope on earth? The temperature dependent orientation errors of the reflector are about $4''$.

(d) How many photons were emitted and how many were received?

(e) How large is the fluctuation of the arrival time, if the quantity (n − 1) with n = refractive index of the air, changes by 1%?

2.6 A radar pulse from the earth to Venus and back takes at the smallest distance between earth and Venus about $\Delta t = 276$ s. How long would it take 30 days later (we assume circular orbits for earth and venus).

2.7 Show that the difference $\varrho = \xi_r - \xi_a$ between real and apparent zenith distance $\xi = (90° - h)$ due to refraction in the earth's atmosphere is given by

$$\varrho = a \cdot (n(0) - 1) \tan \xi_a.$$

2.8 Two stars with the same absolute magnitude $M = 1.0$ have the apparent magnitudes $m_1 = -1.0$ and $m_2 = +2.0$. How large are their distances from the earth?

2.9 A star with radius R has the luminosity L. What is its surface temperature? Numerical example: $R = 7 \cdot 10^8$ m, $L = 10^{27}$ W.

2.10 The star Canopus (α-Carinae) has the absolute visual magnitude $M_V = -4.6$ and the apparent visual magnitude $m_v = -0.71$. What is its distance from the earth?

References

1. H. Karttunen et.al: Fundamental Astronomy (Springer Heidelberg 2016) Peter Duffet-Smith: Practical Astronomy with your calculator (Cambridge University Press 2017)
2. Blaauw, A.; Gum, C.S.; Pawsey, J.L.; Westerhout, G. (1960). The new IAU system of galactic coordinates (1958 revision). Monthly Notices of the Royal Astronomical Society. 121 (2): 123.
3. Jingquan Cheng: The Prinicples of Astronomical Telescopes (Springer Hei delberg 2009)
4. Henry C. King; The History of the Telescope (Dover Books on Astronomy, 2011)
5. R.N. Wilson. Reflecting Telescope Optics I and II (Springer Heidelberg 1996)
6. Ch. Woodhouse: The Astro-Photography Manual (Focal Press 2017)
7. Jacob W.M. Baars et. Al: Radio Telescope Reflectors—Historical Development of Design and Construction. (Springer, Heidelberg 2018)
8. Coude' telescopeIArticle about coude telescope by The Free Dictionary
9. Schmidt–Cassegrain telescope—Wikipedia
10. Jacquot, P.: Speckle interferometry: a review of the principal methods in use for experimental mechanics applications. Strain 44, 57–69 (2008)
11. Active optics—Wikipedia
12. J.W. Hardy: Adaptive Optics for Astronomicaol Telescopes (Oxford JUniv. Press 1998) F.Roddier: Adaptive Optics in Astronomy (Cambridge IUniv. Press 2004)
13. ESO—The European ELT ESO's VLT Takes First Detailed Image of Disc around Young Star". ESO Announcements. Retrieved 17 November 2011
14. Radioteleskop Effelsberg—Wikipedia
15. https://greenbankobservatory.org/science/telescopes/gbt
16. https://en.wikipedia.org/wiki/Arecibo_Telescope
17. https://www.wendyweitours.com/fast-telescope-guizhou.htm
18. Very-long-baseline interferometry—Wikipedia
19. ESO—Interferometry Tutorial Aerial view of the VLTI with tunnels superimposed | ESO
20. ROSAT—Wikipedia
21. Projects | Max Planck Institute for extraterrestrial Physics (mpg.de) X-ray astronomy satellite—Wikipedia "NASA's Explorer Program Satellites". NASA Space Science Data Coordinated Archive. 2014
22. Cortina, J.; for the MAGIC collaboration (2004). "Status and First Results of the MAGIC Telescope". Astrophysics and Space Science. 297 (2005): 245–255
23. H.E.S.S.—The High Energy Stereoscopic System (mpg.de) Search for dark matter annihilation in the dwarf irregular galaxy WLM with H.E.S.S. Collaboration: H. Abdalla et al. Phys. Rev. D 103, 102002 (2021)
24. https://de.wikipedia.org/wiki/Integral_(Satellit)#:~:text=Integral)
25. X-ray astronomy satellite—Wikipedia Introduction to X-ray Astronomy (cam.ac.uk) X-ray telescope—Wikipedia
26. Gravitational wave: Wikipedia https://en.wikipedia.org/wiki/Gravitational_wave#:~:text=Gravitational%20waves
27. Abbott, R.; et al. (29 June 2021). "Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences". The Astrophysical Journal Letters. 915 (1): L5. arXiv:2106.15163. Bibcode:2021ApJ...915L...5A. https://doi.org/10.3847/2041-8213/ac082e. S2CID 235670241.

Measuring Techniques and Detectors

3

Contents

3.1 Measuring Techniques in Astronomy 37
3.2 Measurements of Star Distances 39
3.3 Apparent and Absolute Magnitudes of Stars 46
3.4 Measurements of the Spectral Distribution of Star Radiation 47
Summary 48
Problems 48
References 48

In this chapter, we will treat the different measuring techniques in astronomy, the advantages and limitations and the corresponding detectors, their sensitivity and their spectral response. In particular, we will discuss the different phenomena, which have to be taken into account in order to obtain reliable results.

3.1 Measuring Techniques in Astronomy

In order to keep a telescope always pointing to the same position in the sky, one has not only to compensate for the earth's rotation (see Sect. 2.6) but has to take into account several other effects. One of them is the annual revolution of the earth around the sun, which causes a circulation of the star image around a fixed position, called the *parallax*.

3.1.1 Parallax

Seen from the earth, its circulation around the sun lets the star images perform small ellipses around a line between sun and star (Fig. 3.1). The form of the ellipse depends on the position of the star relative to the ecliptic (orbital plane of the earth). If the radius vector from the star to the sun is perpendicular to the ecliptic, the ellipse is a circle (Fig. 3.1a). If the radius vector to the star lies in the ecliptic, the apparent motion of the star is a sinusoidal motion on a straight line (Fig. 3.1b). For the general case that the radius vector forms the angle α ($0° < \alpha < 90°$) against the ecliptic, the apparent motion of the star occurs on an ellipse with a ratio a/b of large and small semiaxes.

The angle π between the line sun-star and the line earth-star is equal to the angle under which the distance sun-earth (1 AU) appears as seen from the star. It is also called the **parallax of the star** [1].

Since the mean radius of the earth's orbit around the sun is known (it is the astronomical unit 1AU), the parallax π of a star gives its distance r from the sun (Fig. 3.2).

$r = 1\,\text{AU}/\tan\pi \approx 1\,\text{AU}/\pi$.

Example

For $\pi = 1''$ the distance is

$$r = 1\,\text{Parsec} = 1\,\text{pc}$$
$$= 3 \cdot 10^{16}\,\text{m} = 3.26\,\text{lightyears}.$$

◀

The parallax π of a nearby star can be measured, if its apparent position (i.e. the two angles (hour angle and declination) and its variation during one year is compared with the position of a far distant star, with a very small parallax.

Example

The parallax of the closest star P*roxima Centauri* is $\pi = 0.''76 \Rightarrow r = 1.31\,\text{pc} = 4.27\,\text{ly}$. The parallax of the star Betelgeuze (α-Orionis) is $\pi = 0.''00763 \Rightarrow r = 131\,\text{pc} = 427\,\text{ly}$. ◀

Besides this annual parallax π_a also a daily parallax π_d occurs (Fig. 3.3), where the earth diameter is the basis instead of the diameter of the earth's orbit around the sun. It is the angle under which the equator radius appears from the observed object. The daily parallax is therefore much smaller than the annual parallax and can be only applied to much closer objects such as objects in our solar system.

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_3

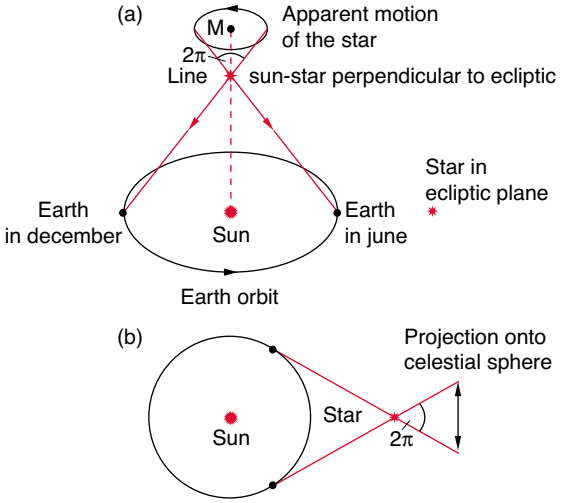


Fig. 3.1 Due to the elliptical orbit of the earth around the sun the stars appear to move on elliptical paths. **a** Direction of view perpendicular to the ecliptic; **b** parallel to the ecliptic

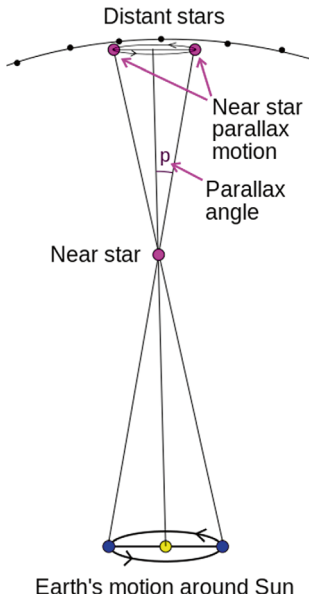


Fig. 3.2 Annual parallax of a near star at the distance $r = 1/\pi$ parsec

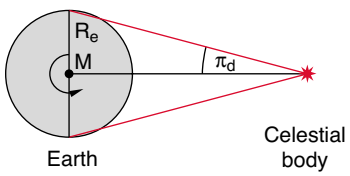


Fig. 3.3 Daily parallax caused by the rotation of the earth. The magnitude of the earth is greatly exaggerated

Example

The daily parallax of the moon is $\pi_d = 57'$, that of the sun is $\pi_d = 8.''79$, that of Neptun is $\pi_d < 0.''3$. ◀

3.1.2 Aberration

Another effect, which shifts the measured position of a star against its true position is the aberration (Fig. 3.4). During the travel time $\Delta t = L/c$ of light through the telescope with length L the earth moves with the velocity $v_E = 30$ km/s around the sun. If the radius vector **r** from the star to the telescope forms the angle ϑ against the velocity $\mathbf{v_E}$ the telescope axis has to be tilted against **r** by the aberration angle

$$\alpha = \frac{v_E}{c} \cdot \sin \vartheta \tag{3.1}$$

in order to assure that the light from the star passes exactly along the telescope axis [2].

The observer therefore sees a celestial object under the angle $(\vartheta + \alpha)$ shifted against its true angle ϑ by the angle α.

The telescope must be aligned against the direction of $\mathbf{v}_E$ by the angle $(\vartheta + \alpha)$. Since the velocity v_E is always the tangent to the earth's orbit, the aberration during the revolution around the sun results in an apparent elliptical path of the observed star image with an eccentricity that depends on the angle ϑ between the radius vector r and the earth velocity v_E. For $\vartheta = 0$ the path is a straight line, for $\vartheta = 90°$ it is a circle. The semi-major axis is for $0 \leq \vartheta \leq 90°$ for all stars equal. It is called *aberration constant* $A = 20.''5 = \bar{v}_E/c$, while the semi-minor axis varies from zero ($\vartheta = 0$) to A ($\vartheta = 90°$).

Besides the annual aberration, there is also a daily aberration due to the earth's rotation, similar to the situation for the parallax.

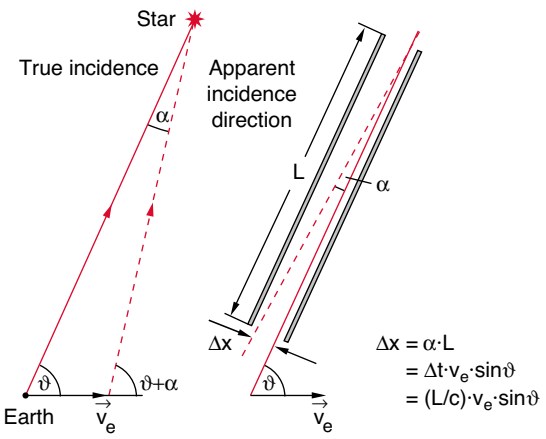


Fig. 3.4 Aberration of light

For an observer on earth at the latitude φ the velocity of the earth's rotation is $v = v_{\text{equator}} \cdot \cos\varphi$. The aberration angle α is then

$$\alpha = \frac{v_{\text{equator}} \cdot \cos\varphi}{c} \cdot \sin\vartheta \tag{3.2}$$

Example

The maximum annual aberration for $\vartheta = 90°$ is $\alpha = v_E/c = 10^{-4}\text{rad} \hat{=} 20.''5$. The daily aberration caused by the rotation of the earth is even at the equator only $a < 0.''3$. ◀

These numerical examples show, that the aberration is for all stars large compared to their parallax. The observed elliptical paths of the star images are a superposition of aberration and parallax.

The essential difference is, however: Stars with a small angular difference but different distances show the same aberration but different parallaxes. In order to distinguish between the two effects, which is necessary for measuring the parallax of a star, one has to choose nearby stars with sufficiently large parallaxes.

3.1.3 Refraction in the Atmosphere

The determination of star positions with earth-bound telescopes meets a further perturbing effect: The light from the star has to pass through the earth's atmosphere where it is refracted due to the radial gradient of the refractive index (Fig. 3.5). The refraction angle $\varrho = \zeta_t - \zeta_a$ is equal to the difference of the true and the apparent zenith distance. As has been derived in Vol. 2 Chap. 9.7 the relation between the diffraction angle ρ and the measured zenith distance is given by

$$\varrho = a(n_0 - 1) \cdot \tan\zeta_a \tag{3.3}$$

where n_0 is the refractive index of the atmosphere at the ground and a is a constant which depends only slightly on ζ_a. The refraction angle depends on the height H of the observation point above sea level, because n_0(H) depends on H.
In Fig. 3.5b the refraction angle ρ is plotted as a function of the true angular height $h = 90^\circ - \zeta$ of a star.

Besides aberration and refraction, which both cause a difference between true and measured positions of stars, the fact that the earth as platform for the observing astronomer is by no means a basis fixed in space. Besides the (not exactly uniform) daily rotation and annual revolution around the sun, which is also not exactly uniform due to gravitational interactions with the moon and other planets, the precession and nutation of the earth's axis cause further changes of the orientation of the earth in space (see Vol. 1 Sect. 3.5.8).

This illustrates that the astronomer has to take into account many effects which makes it difficult to obtain the *true positions* of stars from the measured *apparent positions*. It is therefore admirable how precise modern astronomy can determine true star positions and their variation in time [3].

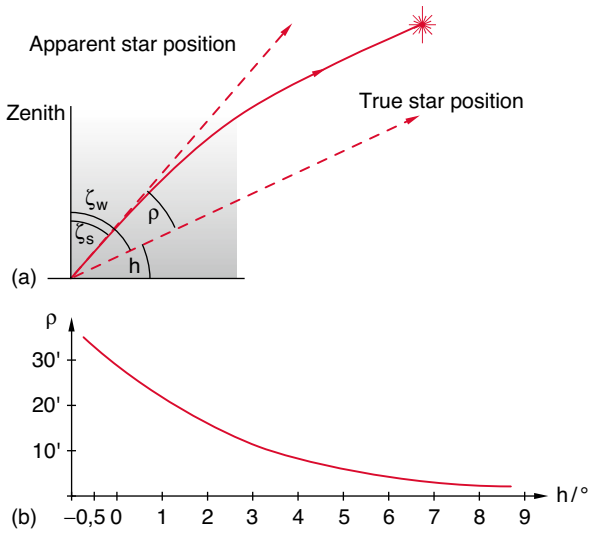


Fig. 3.5 **a** Refraction of star light in the atmosphere. **b** Refraction angle ρ (in arc minutes) as a function of the true height h (in degrees) of a star

3.2 Measurements of Star Distances

Measurements of distances of stars and galaxies are one of the central but also most difficult problems of astronomy. In particular, for far distant objects many effects can influence the accuracy of distance measurements. In order to interfere from the data of an astronomical object measured on earth to the real data, one generally has to know its distance. Therefore, great efforts are undertaken to develop accurate techniques of distance measurements. Often new techniques have corrected the results of earlier measurements [4, 9].

3.2.1 Geometrical Methods

The distances of our nearest celestial objects (Moon, Sun, neighboring planets) can be determined by triangulation (Fig. 3.6). The length 2a of a basis between the points A and B and the two angles α and ß are measured. From Fig. 3.6 we obtain

$$b = a \cdot \tan\alpha \text{and} x = R \cdot \cos\gamma \tag{3.4}$$

With $\alpha + ß = 90° + \gamma \rightarrow \gamma = \alpha + \beta - 90°$.

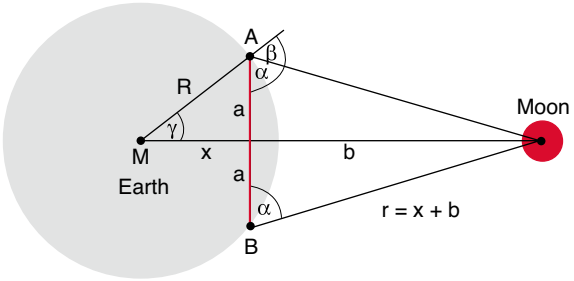


Fig. 3.6 Distance measurements of close celestial objects by triangulation

The distance $r = b + x$ between the centers of moon and earth is then

$$r = b + x = a \cdot \tan\alpha + R \cdot \sin(\alpha + \beta) \tag{3.5}$$

For $r \gg R$ is $b \gg x$ and the first term in (3.5) is much larger than the second term. The accuracy of the distance measurement is therefore essentially given by that of the first term. Since the distance a can be measured very accurately, the determination of the angle α limits the total accuracy.

Example

Trigonometric measurements of the distance to the moon: Here is $r = 380000$ km. With $2a = 1000$ km we get $\tan\alpha \approx r/a = 760 \rightarrow \alpha = 89.98°$. With an accuracy of $0.1''$ for the angle measurement $\rightarrow \delta\alpha = 0.000030°$. The true value of tan α then lies within the range $\tan\ 89.97997 = 6.300920 < \tan a < \tan\ 89.98003 = 6.300958$ $\rightarrow \delta r = 14.4$ km. The relative uncertainty of the distance measurement is then $\delta r/r = 3.8 \cdot 10^{-5}$. ◀

Note: The distance earth-moon can be determined much more accurately by measuring the revolution time of the moon around the earth and applying Kepler's law (see Sect. 4.2).

3.2.2 Time of Flight Measurements with Laser Pulses

With modern laser pulse techniques, where the time of flight of very short laser pulses from a point on earth to a retroreflector on the moon and back is measured with an accuracy of $\Delta t \leq 10^{-10}$ s. This results in an accuracy of 0.01 m (!!) for the distance measurement, which is essentially limited by the uncertainty of the radial progression of the refractive index of the earth's atmosphere and the small number of received photons.

The measuring technique is illustrated in Fig. 3.7.

The beam of a short-pulsed laser is widened by a telescope in order to reduce the diffraction-limited divergence.

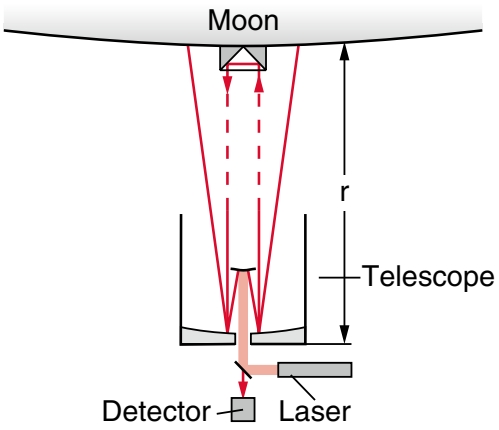


Fig. 3.7 Measurement of the distance of earth-moon by measuring the time of flight of short laser pulses

The light reflected by the retroreflector on the moon (Fig. 3.8) is collected by the same telescope and the time difference Δt between the emitted light ($t_1 = 0$) and the received reflected light ($t_2 = 2r/c$) is measured with an accuracy of $\Delta t < 10^{-10}$ s. The distance to the moon is then

$$r = \frac{1}{2}\Delta t \cdot c.$$

With the accuracy $\Delta t = 10^{-10}$ s, the uncertainty of the distance is $\Delta r = \frac{1}{2}\Delta t \cdot c = 1.5$ cm (!!!), Since the refractive index $n(h)$ of the air is only known within certain limits, the real uncertainty of the distance is somewhat larger [5].

$$\delta r = c \cdot \Delta t + \delta(n-1) \cdot 10^4 \text{ m} \tag{3.6}$$

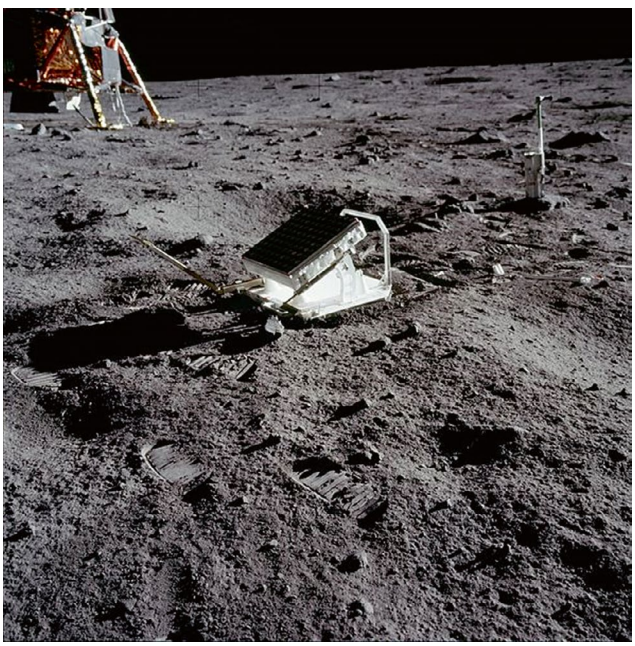


Fig. 3.8 Retroreflector on the moon (Apollo Mission, NASA)

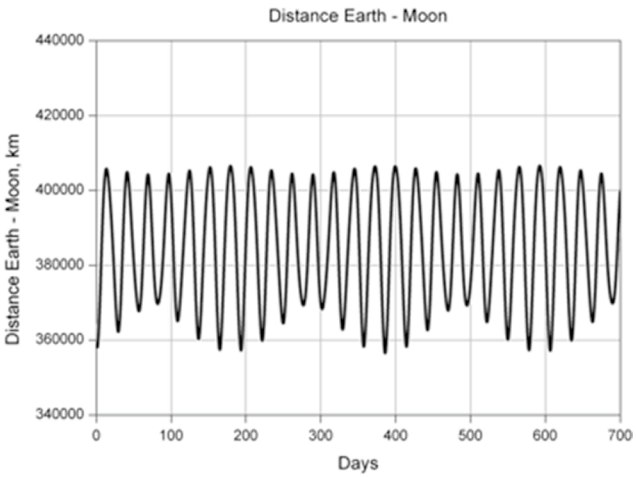


Fig. 3.9 Periodic variations of the distance between the centers of earth and moon (Wikipedia)

where the last term describes the uncertainty of the optical path through the earth's atmosphere because of the limited knowledge of the refractive index $n(h)$.

If this distance is measured from two different points P_1 and P_2 on Earth at different times the annual change of the distance $\overline{P_1P_2}(t)$ (e.g. by the continental drift) can be accurately determined [6].

Measurements over several years proved that the lunar distance increases per year by 3.8 cm, due to friction of the tides on earth and moon. Furthermore, it shows a periodic change (Fig. 3.9), due to the elliptic orbit of the moon around the Earth.

In a similar way, the distances to our nearest planets (Venus and Mars) and to Planetoids (e.g. Eros) can be measured. In these cases, one uses Radar-Pulses in the microwave range instead of Laser pulses in the visible. With this technique, the absolute distances between earth and planet can be obtained for different positions of the planet (Fig. 3.10). The relative distances a_E/a_i of earth and planet to the sun are obtained from the 3. Kepler's law, which states that the squares of the circulation periods are proportional to the 3rd power of the major axes of the elliptical path around the sun

$$\frac{T_E^2}{T_i^2} = \frac{a_E^3}{a_i^3} \tag{3.7}$$

where the index i stands for a planet or planetoid and E for the earth. It is (see Sect. 4.1.2)

$$T_i^2 = \frac{4\pi^2}{G(M_\odot + m_i)} \cdot a_i^3 \tag{3.8}$$

where G is the gravitational constant. From the measured values T_i for planet i and T_E for the earth we obtain $M_\odot >> m_i$

$$\frac{a_i^3}{a_E^3} = \frac{M_\odot + m_i}{M_\odot + m_E}\frac{T_i^2}{T_E^2} = \frac{1 + \frac{m_i}{M_\odot}}{1 + \frac{m_E}{M_\odot}}\frac{T_i^2}{T_E^2} \approx \frac{T_i^2}{T_E^2}. \tag{3.9}$$

Application of the cosine rule on the triangle ESV in Fig. 3.10 gives

$$a_i^2 = a_E^2 + x^2 - 2a_E x \cdot \cos\alpha \tag{3.10}$$

With (3.9) a_i can be expressed by a_E. Measurements of the angle α for different positions of Earth and Venus and measurements of the absolute distances between Earth and Venus for different values of α allow the determination of the major semi-axis a_E of the earth orbit around the sun. These measurements yield the numerical value

$$a_E = 149.597.895\,\text{km}$$

In former times this was the value of the astronomical unit 1AU. Since the accuracy of this value is no longer sufficient for demands of modern astronomy a new definition of the AU was accepted in 1976, which is based on the gravitational law. It reads:

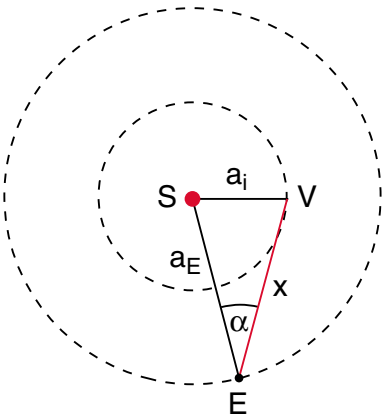


Fig. 3.10 Absolute measurement of the astronomical unit 1 AU

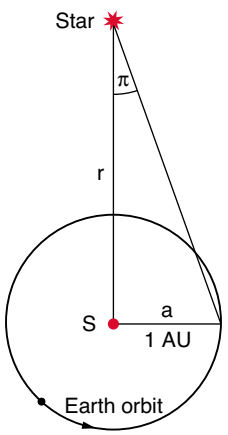


Fig. 3.11 Relation between parallax π and distance r of a star

1AU is the radius of a circular path around the sun, which a hypothetical body with negligibly small mass traverses within 365.2568983 days (Fig. 3.11).

With

$R = G \cdot M_E \cdot T^2/4\pi^2 \rightarrow$

1AU = R(T = 365.2568983 days).

The numerical value accepted today is

$$1\,\text{AU} = 1.49597870700 \cdot 10^{11}\,\text{m} = 149597870.700\,\text{km}. \quad (3.10)$$

The accuracy of this value is only limited by the uncertainty of the gravitational constant G.

This value is only slightly smaller than the major semi-axis of the eliptical path of the earth

$$a_E = 1.000000636\,\text{AU}$$

Note: Since the solar mass $M_\odot$ is very large compared with the mass of the planets ($M_\odot = 3.3 \cdot 10^5\, m_E$) the prefactor in (3.9)

$$\frac{1 + m_i/M_\odot}{1 + m_E/M_\odot} \approx 1$$

can be omitted in most cases. However, for very precise measurements it must be taken into account and one has to know the mass ratios $m_i/M_\odot$ and $m_E/M_\odot$ for an accurate determination of this prefactor in (3.9). For the determination of the absolute masses, the accuracy is limited by the uncertainty of the gravitational constant G (see Sect. 4.2). The mass of the sun can be obtained from the radius r_i and the revolution period T_i of a planet i according to the equation

$$m_i \cdot \omega_i^2 \cdot r_i = m_i, r_i \cdot 4\pi^2/T_i^2 = G \cdot m_i \cdot M_{ik}/r_i^2 \quad (3.11)$$

while the mass of a planet can be obtained from the revolution period of the moon of this planet.

The astronomical unit 1 AU and the astronomical distance unit $1\text{parsec} = 1\text{pc} = r(\pi_s = 1'')$ are used as the basis of measurements of interstellar distances r.

The distance r of a star from the center of the sun is then in units of AU

$$r = \frac{1\,\text{AU}}{\pi_s} \quad (3.12)$$

where π_s is the angle, under which the distance 1 AU is seen from the star.

Example

A star with $\pi_s = 1''$ ==> $1'' = (1/206265)$ rad has the distance $r = 1\,\text{pc} = 206265\,\text{AU} = 3.085677 \cdot 10^{16}\,\text{m} = 3.26$ Ly from the center of our sun. ◀

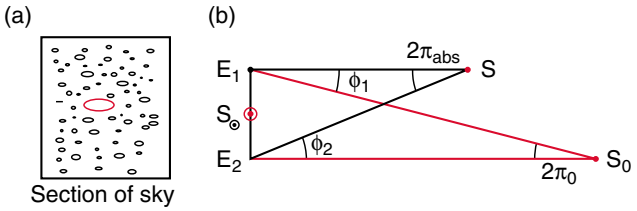


Fig. 3.12 **a** Schematic representation of the analysis of measured parallaxes of close stars (red ellipse) and far distant stars within a sky section. **b** Principle of absolute parallax measurements from two points of the earth orbit by comparison of the parallaxes of far distant stars (after U. Bastian, Astronomie und Raumfahrt **33**, 10 (1996))

3.2.3 Trigonometric Parallax Measurements

The first trigonometric parallax measurement was performed by *Friedrich Wilhelm Bessel* (1784–1846) for the star 61 Cygni. He obtained $\pi_s = 0.''3$.

Until 1990 the parallaxes of about 8000 stars have been measured. For conventional measurements with earth-bound telescopes, the mean uncertainty is for most stars of the same magnitudes as the parallax itself.

All parallaxes measured with earth-bound telescopes give relative values, because the measured parallax of a nearby star is compared with that of far distant stars. Since the absolute values of star positions cannot be measured with the same accuracy as the angular differences of stars with nearly the same angular coordinates (which are close neighbors on the celestial sphere) one takes several photo-plate exposures during a year. In Fig. 3.12 such photo-plate exposures on a plate are schematically illustrated, which shows the parallax of a close star (red ellipse) and of far-away stars. It illustrates the difficulties of obtaining reliable values of parallaxes, because the measured elliptical track on the photo-plate is the sum of the much larger aberration and the smaller parallax.

3.2.4 Astrometric Satellite Gaia

The situation has drastically improved with the astrometric satellites Hipparcos and Gaia. **Gaia** (Global **A**strometric **I**nterferometer for **A**strophysics, Fig. 3.13) was launched in 2013 and brought into a Lissajous path around the Lagrange point L_2 between sun and earth (Fig. 3.14). It measures the position and luminosity of about 1 billion Stars (1% of all stars in our galaxy!) with an unprecedented accuracy. The angular position of stars with a luminosity of $m < 15$ magnitudes (see next section) can be measured with an uncertainty $\delta < 25 \times 10^{-6}$ s of arc. For stars with $m < 20$ the angular uncertainty increases to $(300 \times 10^{-6})''$. For stars with $m < 16$, even their spectra can be measured with a spectrograph on board of Gaia.

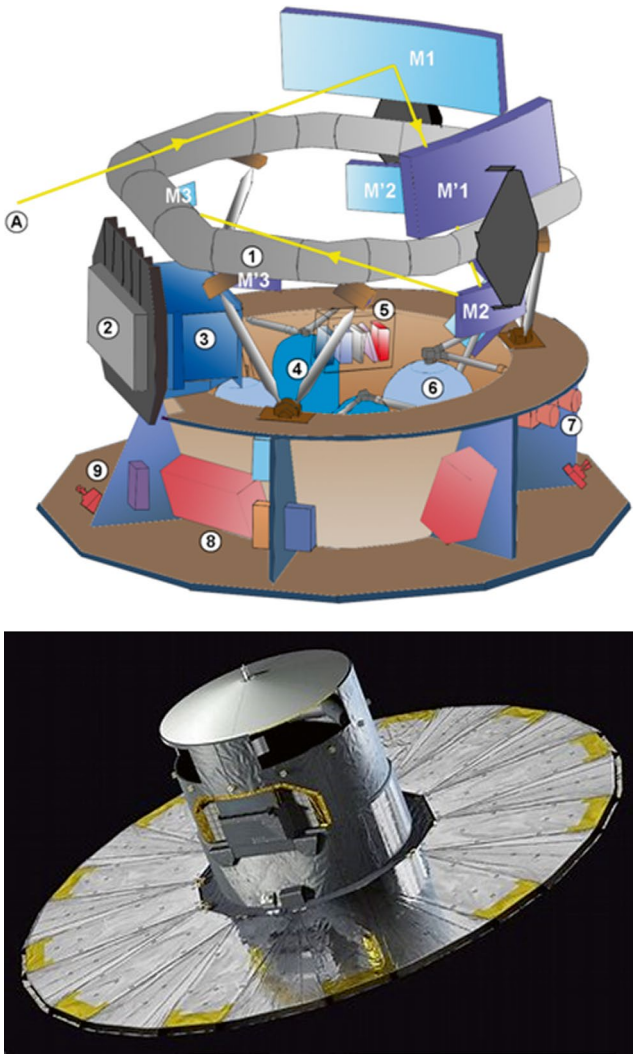


Fig. 3.13 The satellite Gaia for measuring star positions and luminosities (ESA and https://de.wikipedia.org/wiki/Gaia_(Raumsonde))

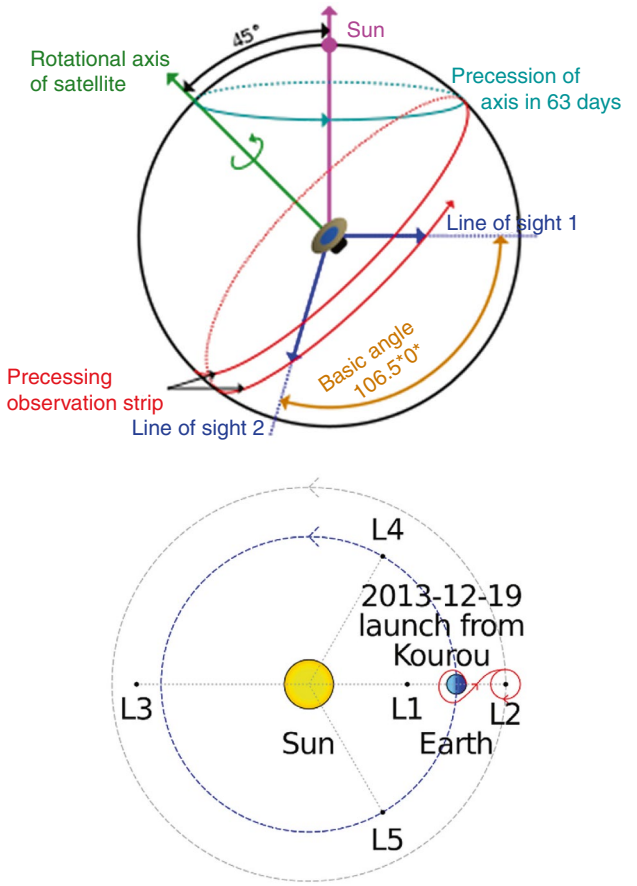


Fig. 3.14 Position and Alignment of Gaia in the Lagrange-point L_2 (green: rotational axis of satellite, red: progressive observations stripe, blue: line of sight, bright blue: precession of axis within 63 days) (By C. Mglee—Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=30257831)

How is this high accuracy achieved?
Gaia carries two mirror telescopes which are directed against different parts of the sky with widely separated angular coordinates. Both telescopes focus the incident light onto a common focal plane where 106 very sensitive CCD-detectors are located which detect the spectral range from the near UV (330 nm) to the near infrared (1050 nm). Each of these detectors consists of 4500×1966 pixels which gives altogether nearly 1 billion pixels, which makes it the most sensitive camera in space with the highest spatial resolution.

63 of these CCD detectors are located in a separate field arranged as a 7×9 pattern which measures the same stars several times at different times i.e. different positions of GAIA, in order to determine their parallax and therefore their distance from the satellite. The analysis of the data after many measurements allows an accuracy of 9×10^{-6} arcsec. Multiple measurements of star positions during the operation time of Gaia also provide the determination of the angular velocities of stars (Fig. 3.15).

Two photometers with 14 CCD detectors measure the photometric luminosity of stars and give information about their surface temperature.

The spectrograph measures some prominent lines of atomic elements, which gives information about the chemical composition of the star atmosphere, the spectral classification of the star (see Sect. 5.2) and its radial velocity from the Doppler-shift of the spectral lines.

More detailed information about Gaia and an extensive literature can be found in the Wikipedia article on Gaia [7].

In Fig. 3.16 the principle of parallax measurements with the older satellite HIPPARCOS is illustrated. The optics in HIPPARCOS contain a two-section mirror which images

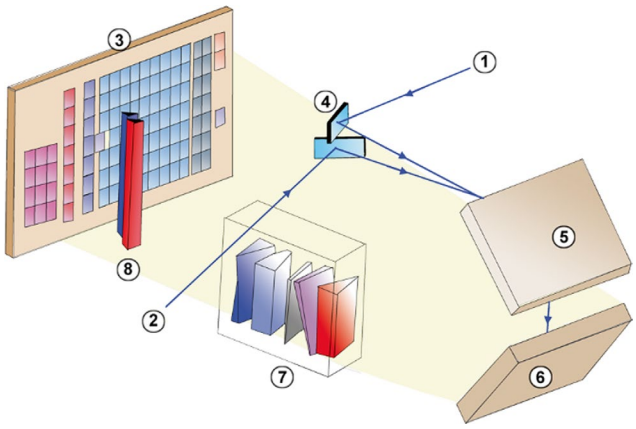


Fig. 3.15 Measuring scheme of Gaia. 1 = incident light from mirror M1 in Fig. 3.13, 2 = incident light from mirror M3; 3 = focal plane, 4 = mirrors M4 and M′4; 5 = mirror M5; 6 = mirror M6, (ESA)

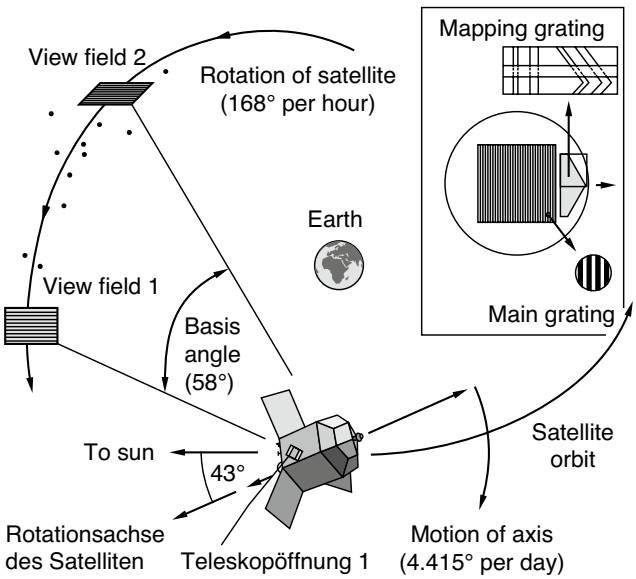


Fig. 3.16 Principle of position determination and measurements of parallaxes with the astrometric satellite HIPPACOS (after U. Bastian, Astronomie und Raumfahrt **33** 1026 1996)

two stars with an angular distance of 58° onto the object mirror of the telescope. The two fields of view of the two stars are exactly superimposed in the focal plane of the telescope. This has the advantage that stars with a very small angular distance of about 1° and with a large angular distance of 58° can be simultaneously measured. The satellite rotates within 2 h around a full circle, i.e. it covers within 2 h a full circle of 360° on the sky.

These very accurate measurements of GAIA have provided benchmarks for other distance measurements and allow the calibration of the Cepheid-method.

3.2.5 Other Methods for Measurements of Star Distances

For measurements of distances of stars in other Galaxies, the conventional trigonometric methods are not sufficiently accurate and other indirect methods are demanded [9]. The new results of GAIA can be used to calibrate these indirect methods. One of them relies on stars with periodic pulsating radius, causing periodic luminosity variations, such as the Cepheids (see Sect. 3.5.8). The term "Cepheid" originates from the pulsating star *delta cepheid* in the constellation *Cepheus*. The pulsation period T is directly related to the absolute luminosity. It was found that for the periods of two cepheids, the relation holds

$$T_1/T_2 = 10^{(m1-m2)/a} \tag{3.13}$$

where m_i is the apparent magnitudes and a is a constant which depends on the special type of Cepheids. Since all cepheids in a globular cluster have the same distance the apparent magnitudes m of these stars are proportional to the absolute magnitudes M.

Measuring the apparent luminosity and the pulsation period one can determine the distance of the star [9]. This method anticipates a physical model of the pulsations and assumes that all Cepheids obey the same relation (Fig. 3.17).

Another method uses binary stars, where their angular distance and their revolution period around their common center of mass are measured (see Sect. 5.1). Using Kepler's laws one can determine the absolute separation of the two stars and therefore from the measured angular separation the absolute distance to the binary system.

For objects at very large distances (for example far-away galaxies) individual stars cannot be separated and a special Doppler-technique is used for determining their distances. It turns out that all galaxies have relative velocities v against the earth which are proportional to their distance r. This was first observed by *Edmund Hubble*, who formulated the relation

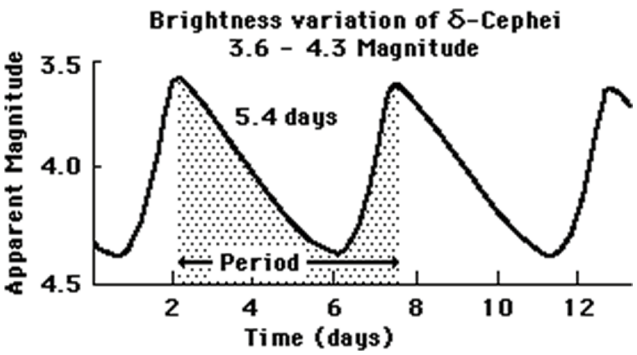


Fig. 3.17 Brightness variation of the variable star δ. Cephei (Wikipedia)

$$v = H \cdot r \tag{6.14}$$

where H ≈ 70 km/(s Mpc) is the Hubble parameter (see Sect. 6.2). This means that a galaxy at a distance of $r = 10^6$ pc recedes from the earth with the velocity 70 km/s. This leads to a Doppler-shift of the spectral lines emitted from the galaxy. A line with the frequency ν_0 is shifted to the lower frequency

$$\nu = \nu_0(1 - v/c) \tag{3.15}$$

The wavelength λ_0 is then red shifted to

$$\lambda = \lambda_0(1 + v/c) \tag{3.15}$$

Measuring the Doppler-shift allows the determination of the distance r [10].

3.2.6 Supernovae as Distance Candles

Recently a new method of measurements of very large distances has been applied which can also determine the distance changes with time and can solve the question of whether expansion of our universe is accelerated or slowing down. It is based on measurements of supernovae of type Ia (see Sect. 5.x), which belong to the brightest supernova explosions in the universe. Since the total energy released in such a supernova Ia explosion is known from reliable model calculations, these supernovae can be used as good standard candles for distance determinations [8].

A supernova Ia occurs when a binary system of a white dwarf and another star (e.g. a red giant) comes so close together (below the Roche limit) that the white dwarf can attract matter from its companion. It therefore not only increases its mass but also gains energy from the fast-impinging matter onto its surface. Therefore, its temperature increases. If it reaches the ignition temperature of carbon fusion (see Sect. 3.4.x) processes start within a few seconds. They further increase the temperature up to many billion degrees and accelerate the fusion processes, which proceed to the production of iron and nickel. The energy released during these processes is about 10^{44} J, which is more than sufficient to let the white dwarf explode. Since the mass of the white dwarf is known, the explosion energy can be calculated. Most of it is converted into radiation. The sudden increase of the luminosity by many orders of magnitude makes the supernova appear much brighter than the luminosity of the whole galaxy. The typical absolute visual magnitude is—19.3 M (see next section).

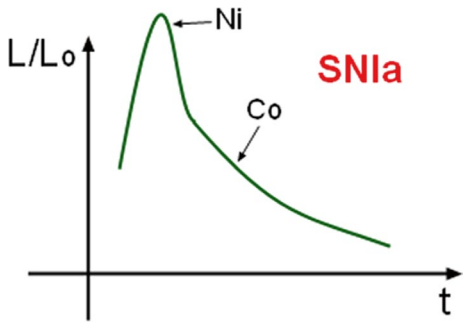


Fig. 3.18 Typical light curve of a supernova Ia (Wikipedia)

Type Ia supernovae have a characteristic light curve (Fig. 3.18), The explosion energy causes a steep increase in the luminosity The photons emitted during the radioactive decay of nickel 56 which decays into cobalt and later into iron ^{56}Fe are responsible for the decreasing part of the light curve.

Measurements of such light curves $I(t)$ have shown, that all supernovae Ia follow the same mechanism with the same exploding mass. Therefore, all SNIa should have the same absolute luminosity. Therefore, measurements of the apparent luminosity $L_{ap} \propto L_{abs}/r^2$ give their distance r.

A very precise method of measuring large distances ($r > 1$ Mpc) is based on measuring the angular diameter $\Delta\alpha = D/r$ of the shock-front emitted during a supernova Ia explosion, which spreads out as a bright spherical shell around the explosion center. It appears as seen from the earth, as a luminous ring with a diameter $D = 2\int v(t)dt$ which depends on the velocity $v(t)$ and the time after the explosion.

Meanwhile, there are very precise models of supernova Ia, which can compute from the measured light curve the velocity $v(t)$ of the shock-front. From the measured angular diameter $\Delta\alpha$ and the computed diameter D the distance $r = D/\Delta\alpha$ can be obtained.

After having determined the distances r of many supernovae Ia at different distances r *Saul Permutter* (US), *Brian Schmidt* (Australia) and *Adam Ries* (US) could prove that the universe expands accelerated. They got for this discovery the Nobel Prize 2011.

3.3 Apparent and Absolute Magnitudes of Stars

The visual luminosity (visual apparent magnitude m of a star appearing to the observer on Earth), is the energy flux $\Phi\left(1\,\mathrm{W/m^2}\right)$ through the eye of the observer, resp. through the aperture of the telescope. The signal received by the

retina cells of the eye and transferred to the brain is, however. not proportional to Φ but rather to log Φ (**Weber-Fechner's law**). One therefore defines the quantity

$$m = -2.5 \cdot \log_{10} \frac{\Phi}{\Phi_0}. \tag{3.16}$$

as the apparent magnitude of a star. The prefactor 2.5 is chosen such, that the magnitude of observed stars agrees with the early classification by the famous astronomer ***Hipparcos*** (≈150 BC) who divided the visible stars into 6 magnitude classes, where the brightest stars were attributed to class 1 the faintest barely visible stars to class 6.

Stars with the energy flux Φ_0 have, according to (3.16) the magnitude $m=0$, and stars with $\Phi = 100\Phi_0$ have the magnitude $m=-5$. **Magnitudes are dimensionless quantities**! In order to illustrate that a number should indicate the magnitude of a star, one writes instead of $m=5$ often $5^{\,\mathrm{m}}$ or 5 mag.

Although this modern definition of the magnitudes of stars follows the old classification of Hipparcos, it extends to both sides far beyond the magnitudes of stars visible with the naked eye, because the observation with telescopes and the use of very sensitive detectors allow the detection and measurement of fare more stars and other celestial objects, such as galaxies or quasars at very large distances and on the high intensity side brilliant events such as supernovae.

Note: The smaller the magnitude m is, the brighter the star. A difference in magnitudes of $\Delta m = 5$ corresponds to a luminosity ratio of 100. For $\Delta m = 1$ is $\Delta\Phi = 2.5$.

Examples

Polar star: $m = +2.^{\mathrm{m}}12$, Sirius: $m = -1.^{\mathrm{m}}5$, Venus: $m_{\max} = -4.^{\mathrm{m}}4$, Full moon: $m = -12.^{\mathrm{m}}5$, sun: $m = -26.^{\mathrm{m}}8$. ◀

The faintest stars, barely visible with the naked eye have the magnitude $m = +5$ to $+6$. With large telescopes, the smallest still detectable apparent luminosity is about $m = +25$.

The magnitudes defined by (3.16) are the **apparent magnitudes**, because they state the brightness as it appeared to the observer on earth. Since the energy flux generally depends on the wavelength of the radiation, the measured signal depends on the spectral sensitivity of the detector. The same star will therefore have different magnitudes, when measured in different spectral ranges. This is why one distinguishes between **visual magnitude** (observed with the eye) **bolometric magnitude** (the spectrally integrated energy flux is measured with a bolometer with a sensitivity independent of the spectral range) or the **spectral magnitude**, measured in a selected spectral range

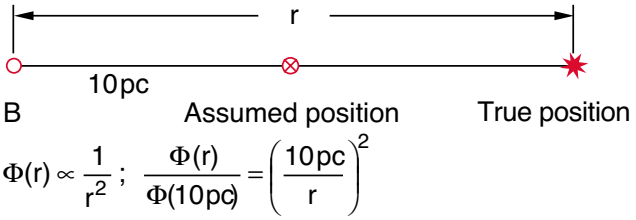


Fig. 3.19 Definition of the absolute magnitude of star radiation

filtered through a spectrometer or through special spectral filters [7].

The energy flux density, measured for a star at a distance r which is large compared to its diameter, decreases with $1/r^2$ (Fig. 3.19). Stars at large distances therefore appear less bright than stars with the same absolute brightness but at smaller distances.

In order to compare the real brightness of different stars the absolute magnitude M is defined as the apparent magnitude m, the star would have at a distance of $r = 10$ pc.

The corresponding radiation flux densities (intensities) are related by

$$\frac{\Phi(r)}{\Phi(10\ \mathrm{pc})} = \left(\frac{10\ \mathrm{pc}}{r}\right)^2 \tag{3.17}$$

This gives the difference between apparent and absolute magnitudes as

$$\begin{aligned} m - M &= -2.5 \cdot \log_{10} \frac{\Phi(r)}{\Phi(10\,\mathrm{pc})} \\ &= -2.5 \cdot \log_{10} \left(\frac{10\,\mathrm{pc}}{r}\right)^2 \\ m - M &= +5 \cdot \log_{10}(r/10\,\mathrm{pc}) \end{aligned} \tag{3.18}$$

where the distance r is given in pc.

Note: Apparent magnitudes are denoted by the small letter m and absolute magnitudes by the capital letter M.

Example

The absolute visual magnitude of our sun is according to (3.17) with $r = 1.5 \cdot 10^{11}$ m, $1\ \mathrm{pc} = 3 \cdot 10^{16}$ m and $m = -26.8$

$$M = m - 5 \cdot \log_{10} \frac{1.5 \cdot 10^{11}\ \mathrm{m}}{3.1 \cdot 10^{17}\ \mathrm{m}}$$

$$= +4.8.$$ ◀

The absolute magnitude M of a star is related to its *luminosity*

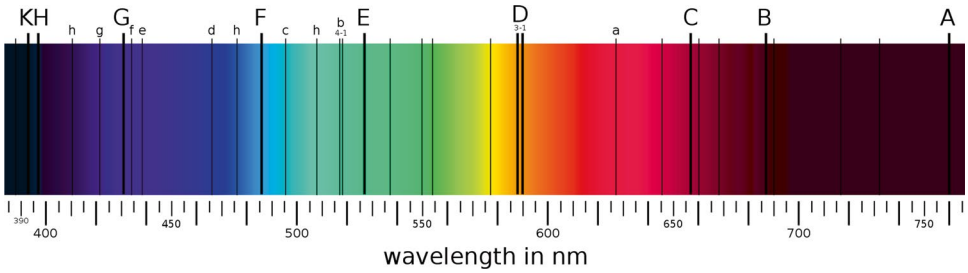


Fig. 3.20 Fraunhofer lines in the sun spectrum (Wikipedia)

$$L = 4\pi r^2 \cdot \Phi \tag{3.19}$$

which is the total radiation power emitted into the solid angle 4π.

Comparing the luminosity L of a star with that of the sun, (which is known, see Sect. 4.6), one obtains the absolute bolometric magnitudes

$$M_{\text{bol}}^{\text{star}} - M_{\text{bol}}^{\odot} = -2.5 \cdot \log_{10} \frac{L}{L_{\odot}}. \tag{3.20}$$

The absolute bolometric magnitude of our sun is with $M_{\text{bol}}^{\odot} = 4.72$ slightly larger than the absolute visual magnitude, because the sun emits radiation into the UV and the infrared, not visible to the eye, although the maximum of the sun radiation lies in the visible range.

The luminosity L of the sun can be determined from the solar constant $\text{SC} = 1367\,\text{W/m}^2$ which falls onto 1 m^2 outside the earth's atmosphere (see Vol. 2 Sect. 12.6). This gives

$$L_{\odot} = 3.9 \cdot 10^{26}\,\text{W} \tag{3.21}$$

With reliable models of a star, one can calculate the absolute magnitude M of a star. From the measured apparent magnitude m one can in principle determine the distance r to the star, using (3.18). However, often absorbing interstellar matter (dust and molecular clouds) partly absorbs the star radiation and also shifts its spectral distribution. This makes distance determinations with this method unreliable (see Sect. 6.5).

Another method of distance measurements of far galaxies is based on Hubble's law.

$$v = H \cdot r \geq r = v/H.$$

between the escape velocity v of a galaxy and its distance r. The velocity v can be determined from the Doppler-shift of the spectral lines. The accuracy is limited by the uncertainty of the Hubble parameter H and by other effects which cause a wavelength shift of the light passing through the space to reach the Earth. Scattering and absorption by the interstellar and intergalactic matter causes a red shift of the transmitted light which overlaps the Doppler-shift. It is often difficult to separate the two effects.

3.4 Measurements of the Spectral Distribution of Star Radiation

The spectral distribution of the star radiation depends on its surface temperature and its chemical composition. Its measurement therefore allows the determination of the surface temperature of a star. Besides the continuous spectral distribution, spectral emission lines, emitted from excited atoms, molecules or ions, appear in the outer star atmosphere. From the intensity and the spectral width of these lines, the abundance of different chemical elements can be inferred as the temperature of the region from where the lines are emitted. Furthermore, absorption lines appear (see Fig. 3.20) which correspond to the radiation emitted by species in hot deeper layers of the star atmosphere and which are absorbed in the outer cooler layers.

The combination of measurements of the continuous and the line spectrum allows a better estimation of the absorption and spectral shifts of the star radiation by interstellar matter.

For the measurements of such star spectra, telescopes with combined spectrographs are used (Coudé-telescope, see Fig. 2.23). If a higher spectral resolution is demanded often interferometers or spectrographs with Echelle gratings are applied. The Echelle devices are grating spectrographs, operated in a high diffraction order m. Their spectral resolution is $\lambda/\Delta\lambda = m \cdot N$ where N is the number of illuminated grooves and m is the diffraction order.

Example

$\text{m} = 5,\ \text{N} = 10^5 \rightarrow \lambda/\Delta\lambda = 500\,000.$ ◀

Since for earth-bound telescopes the star radiation passes through the earth's atmosphere, the measured star spectrum is superimposed by absorption lines of the molecules in the earth's atmosphere, which have to be subtracted.

If spectra of non-self-luminous bodies (moon, asteroids planets) are measured, the measured spectrum represents the sun radiation reflected or scattered by the body. Its spectral distribution is therefore given by the radiation $I_0(\lambda)$, incident

onto the body, times its wavelength-dependent reflectivity $R(\lambda)$

$$I_{\mathrm{B}}(\lambda) = I_0(\lambda) \cdot R(\lambda) \tag{3.22}$$

Knowing the spectral distribution of the incident radiation, the measurement of $I_B(\lambda)$ yields the reflectivity $R(\lambda)$ of the celestial body, which gives information on the density, the composition and the depth of the reflecting layer.

Summary

- The position-Astronomy (Astrometry) develops techniques for a precise determination of star positions (angular coordinates and distances) and their variation in time.
- Since the observer on earth moves with the earth around the sun, the apparent positions of stars seem to move on elliptical paths on the celestial sphere during the year. The major axis of these ellipses is caused by the aberration ($\approx 20''$ and the parallax ($<1''$).
- The parallax $\pi = 1/r(\mathrm{pc})$ of a star depends on its distance r and decreases with increasing distance. It can be described as the angle under which the earth-sun distance (1 AU) appears as seen from the star. The distance r of a star equals the inverse of the parallax $r = 1/\pi$. 1 parsec (1pc) is the distance r for which $\pi = 1''$.
- The refraction in the earth's atmosphere bends the light rays and causes the measured height of a star to appear too high. Aberration and refraction have to be taken into account in order to determine correctly the true position of a star.
- The only direct techniques of measuring star distances are based.

(1) on time of flight measurements of electromagnetic pulses and are only possible for nearby objects (moon and neighboring planets).
(2) on triangulation methods (parallax measurements): With the satellite Gaia the accuracy of such measurements has increased by several orders of magnitude and has reached uncertainties of 10^{-5} s of arc.
(3) The apparent magnitude m of a star is a logarithmic measure of its intensity observed on Earth. It is defined as

$$m = -2.5 \cdot \log_{10} \frac{\Phi}{\Phi_0},$$

where Φ_0 is the intensity received from a star with magnitude m = 0.

- The absolute magnitude M of a star is defined as its apparent magnitude m(10 pc) if the star would be set at a distance of 10 pc.
- The bolometric magnitude is a measure of the integrated radiation flux density $\Phi = \int_0^\infty \Phi_\lambda(\lambda)\mathrm{d}\lambda$.
- Measurements of the spectral distribution of the star radiation yield information on the surface temperature of a star and its chemical composition.

Problems

3.1 A laser pulse (peak power $P = 1\,\mathrm{GW}$, duration $\tau = 150\,\mathrm{ps}$, wavelength $\lambda = 500\,\mathrm{nm}$) is sent through a telescope ($D = 0.75\,\mathrm{m}$) to the moon.
(a) What is the diameter of the diffraction limited laser spot on the moon under ideal conditions?
(b) The real beam divergence due to diffraction and air turbulence (seeing) is $\alpha = 3''$. Which fraction of the emitted photons reach the reflector on the moon ($0.4 \times 0.4\ \mathrm{m}^2$). Which fraction reaches the telescope on earth, if the reflector is assumed to be ideal and the angle of the reflected light is diffraction limited?
(c) How many photons were emitted from the telescope and how many are again received?
(d) How large is the change of the arrival time if the refractive index n of the atmosphere changes by $\delta(n-1) = 1\%$?

3.2 A radar pulse, sent from the earth to Venus at the time of the minimum distance to Venus takes for the round trip the time $\Delta t = 276\,\mathrm{s}$. How long would it take 30 days later (assume circular paths for Earth and Venus)?

3.3 Show, that the difference $\varrho = \xi_\mathrm{t} - \xi_\mathrm{a}$ between the true and apparent zenith distance of a star due to the refraction of light in the earth's atmosphere, is given by

$$\varrho = a(n(0) - 1) \cdot \tan\xi_\mathrm{a}$$

where $n(0)$ is the refractive index of the air at the heights $z = 0$.

3.4 Two stars with the same absolute magnitude $M = 1.0$ have the apparent magnitudes $m_1 = -1.0$ and $m_2 = +2.0$. What are their distances from the earth?

3.5 A star with radius R has the luminosity L. What is its surface temperature? Numerical example: $L = 10^{27}\,\mathrm{W}$; $R = 7 \cdot 10^8\,\mathrm{m}$.

3.6 The star Canopus (α-Carinae) has the absolute magnitude $M_\mathrm{V} = -4.6$ and the apparent visual magnitude $m_\mathrm{v} = -0.71$. What is its distance from the earth?

References

1. Parallax - Wikipedia
2. Aberration (Astronomie) – Wikipedia
3. Atmospheric Refraction | British Astronomical Association (britastro.org)
4. The Distances to Nearby Stars | Astronomy 801: Planets, Stars, Galaxies, and the Universe (psu.edu) https://www.scienceabc.com/nature/universe/how-do-you-measure-the-distance-to-a-star.htm Cosmic distance ladder - Wikipedia ESA Gaia spacecraft summary". ESA. 20 May 2011 "Extended operations confirmed for science missions". ESA. 13 October 2020 https://de.wikipedia.org/wiki/Gaia_(Raumsonde)#:~:text=Gaia%20ist%20ein%20Weltraumteleskop%20der%20Europ%C3%A4ischen%20
5. Lunar Laser Ranging experiment - Wikipedia. https://en.wikipedia.org/wiki/Lunar_distance_(astronomy)
6. ERIC - ED216932 - Measuring Continental Drift: The Laser Ranging Experiment. Crustal Evolution Education Project. Teacher's Guide [and] Student Investigation, 1979
7. https://earthsky.org/astronomy-essentials/what-is-stellar-magnitude/#:~:text=The%20magnitude%20scale%20da Magnitude in Astronomy - Meaning, Scale, Distance Modulus and FAQs (vedantu.com)
8. S. Dhawan, S.W. Jha and B. Leibundgut: Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles https://astrobites.org/2017/11/02/measuring-h0-with-sn-as-near-infrared-standard-candles Type II-P Supernovae as a New Standard Candle - Universe Today
9. Measuring the Distance to Nearby Galaxies (wku.edu)
10. Freedman, W.L. (2000). "The Hubble constant and the expansion age of the Universe". Physics Reports. **333** (1): 13–31

Our Solar System

4

Contents

4.1 Introduction 51
4.2 Kepler's Laws of Planetary Motion 52
4.3 The Inner Planets and Their Moons 63
4.4 The Outer Planets 74
4.5 Small Celestial Bodies in the Solar System 82
4.6 The Sun as Stationary Star 92
4.7 The Active Sun 111
Summary 117
Problems 119
References 119

4.1 Introduction

Our solar system consists of the following components:

- A central star, our Sun, which carries more than 99% of the total mass of the system.
- The four inner planets Mercury, Venus Earth and Mars, which consist mainly of hard rock.
- The four outer planets Jupiter, Saturn, Uranus and Neptune, which are gaseous planets.
- More than 170 moons of these planets.
- Several Dwarf planets such as Pluto, Ceres Eris, Haumea.
- Many small planetoids.
- Comets.
- Many rock particles, dust and microparticles which move on elliptical orbits either around the Sun or in rings around the four outer planets.

The name "*planet*" comes from the Greek name ($\pi\lambda\alpha\nu\eta\zeta$ = wander about). They were in former times named "*wandering star*" in contrast to the "fixed stars", which were assumed to stay in a fixed position.

The International Astronomical Union IAU has defined in 2006 the generally accepted **definition of a planet**:

▶ A planet in our solar system is a celestial body, which

(a) orbits around the Sun,
(b) has sufficient mass to be in a hydrostatic equilibrium due to its self-gravitation and has become a sphere,
(c) has cleaned its surrounding from dust and microparticles due to its gravitational attraction.

The demand (c) does not apply to Pluto (Fig. 4.2), therefore Pluto was downgraded 2006 by the IAU from its status as *planet* to that of a *dwarf-planet* (see Sect. 4.4.1).

There are several more dwarf planets such as *Ceres* in the planetoid belt between Mars and Jupiter and *Haumea*, *Makemake* and *Eris* in the Kuiper Belt beyond the Neptune orbit.

Meanwhile more than 4.400 extrasolar planets have been found which orbit around other stars (see Sect. 8.2).

Much information about our solar system has been collected over many centuries by Earth-bound observers. New essential details about the structure of the planets and their moons, the ring systems around the four outer planets and the composition of comets have been obtained during recent years by space probes (Voyager, Viking, Venera, Helios, Giotto etc.) and by the crewed excursion to the Earth moon [1].

Also, more details about the structure of the Sun surface and dynamic processes changing it (sun oscillations, eruptions, protuberances, granules and magnetic fields) could be investigated with satellites inspecting the Sun in the infrared, ultraviolet and X-ray spectral regions with sufficient time resolution to detect time-dependent processes on the Sun surface. Many space missions have furthermore provided us with a much more detailed knowledge of the chemical composition of planetary atmospheres and have promoted our understanding of the formation of our solar system.

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_4

Table 4.1 Orbital data of the eight planets with the Earth moon for comparison

Name	Symbol	Mayor semi-axis a			Revolution period T	Mean revolution velocity	Numerical eccentricity e	Inclination of orbital plane i	Minimum and maximum distance from Earth	
		In AU	In 10^6 km	Light travel time t		In km s-1			In AU	In AU
Mercury	☿	0.39	57.9	3.2 min	88d	47.9	0.206	7.0°	0.53	1.47
Venus	♀	0.72	108.2	6.0 min	225d	35.0	0.007	3.4°	0.27	1.73
Earth	♁	1.00	149.6	8.3 min	1.00 a	29.8	0.017	–	–	–
Mars	♂	1.52	227.9	12.7 min	1.9 a	24.1	0.093	1.8°	0.38	2.67
Jupiter	♃	5.20	778, 3	43, 2 min	11.9 a	13.1	0.048	1.3°	3.93	6.46
Saturn	♄	9.54	1427	1.3 h	29.46 a	9.6	0.056	2.5°	7.97	11.08
Uranus	♅	19.18	2870	2.7 h	84 a	6.8	0.047	0.8°	17.31	21.12
Neptune	♆	30.06	4496	4.2 h	165 a	5.4	0.009	1.8°	28.80	31.33
Earth moon	☾	0.00257	0.384	1.3 s	27.32 d	1.02	0.055	5.1°	356.410 km	406.740 km

The analysis of all such observations has yielded the facts about our solar system compiled in Table 4.1. Besides the eight planets named in this table, there are a huge number (>10^5) of small celestial bodies, called planetoids and asteroids in the belt between Mars and Jupiter. It is estimated that in the *Kuiper belt* beyond the orbit of Neptune, are besides numerous small bodies at least 10000 objects with diameters of more than 100 km.

In the Oort's Cloud at the fringe of our solar system, about 50 000 AU (≈ 1 ly) away from the Sun a further range of many small bodies is assumed. These bodies are only weakly bound by gravitation of the Sun. They can be therefore strongly affected by other stars passing close to the solar system, which either throws them out of the solar system or bends their path towards the Sun. In this case, they become long period comets that can pass through the inner part of the solar system. In Fig. 4.1 the inner and outer part of our solar system is schematically depicted.

4.2 Kepler's Laws of Planetary Motion

The heliocentric model of our solar system allows a much more simple description of the planetary motions than the geocentric system of Ptolemy which needed elaborate epicyclic orbits to approximately explain the observations, which were mainly performed on our nearest neighbor planet Mars. The heliocentric model can explain the apparent motion of Mars with looplike orbits quite easily without reverting to the epicyclic assumptions of Ptolemy (Fig. 4.3). The loops come about because the observer on Earth moves faster around the Sun than Mars.

4.2.1 Elliptical Planetary Orbits and Kepler's First Law

Due to the gravitational force between Sun (Mass $M_\odot$) and a planet (mass m)

$$\mathbf{F} = -G \cdot \frac{m \cdot M_\odot}{r^2} \hat{\mathbf{r}} \tag{4.1}$$

the planet moves on an elliptical orbit around the center of gravity of planet and Sun, where the Sun is located in one of the focal points. This can be seen from the following reasoning:

The equations of motion are

$$m \ddot{\boldsymbol{r}}_\mathrm{S} = -G \cdot \frac{m \cdot M_\odot}{r^2} \hat{\mathbf{r}} \tag{4.2a}$$

$$M_\odot \ddot{\boldsymbol{r}}_1 = +G \cdot \frac{m \cdot M_\odot}{r^2} \hat{\mathbf{r}} \tag{4.2b}$$

with $\boldsymbol{r} = \boldsymbol{r}_\mathrm{S} - \boldsymbol{r}_1$ (Note that $\mathbf{r}_1$ and $\mathbf{r}_\mathrm{s}$ are antiparallel which means that for their amounts is $r = r_1 + r_\mathrm{s}$ (Fig. 4.4)) and

$$\boldsymbol{r}_1 = -\boldsymbol{r}_\mathrm{S} \cdot m/M_\odot.$$

The subtraction of (4.2b) from (4.2a) gives the equation of motion of the planet in relative coordinates

$$m \ddot{\boldsymbol{r}} = -G \frac{m \cdot (M_\odot + m)}{r^2} \hat{\mathbf{r}}, \tag{4.2c}$$

which differs from (4.1) only by the factor $(M_\odot + m)/M_\odot = (1 + m/M_\odot)$. The integration of (4.2c) gives an elliptical orbit

$$r(t) = \frac{a \cdot (1 + m/M_\odot)\left(1 - \varepsilon^2\right)}{1 + \varepsilon \cos\varphi(t)}. \tag{4.2d}$$

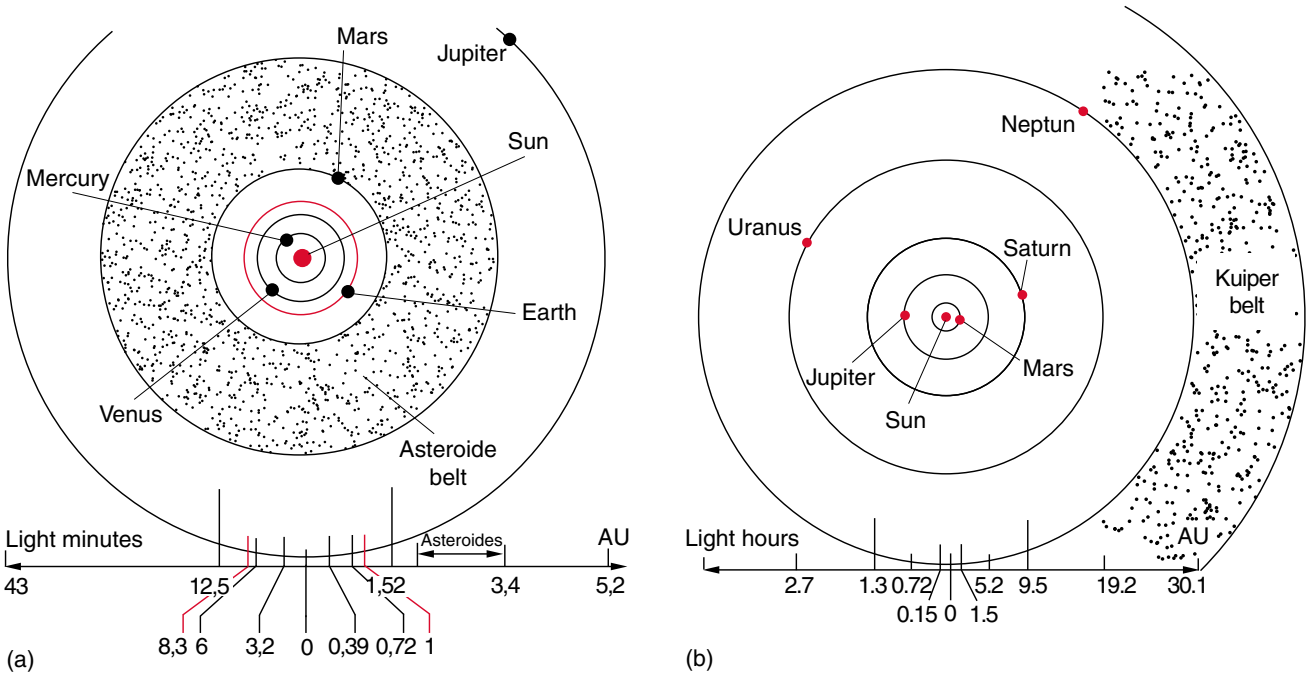


Fig. 4.1 **a** Orbits of the inner planets and the asteroid belt, true to scale, observed from a point above the ecliptic. The orbit of Jupiter is drawn for scale comparison. **b** Orbits of the outer planets. Here the orbit of Mars is included for scale comparison

Fig. 4.2 The dwarf-planet Pluto (false color image) (NASA)

$$r_s(t) = \frac{a \cdot \left(1 - \varepsilon^2\right)}{1 + \varepsilon \cos\varphi(t)} \tag{4.3}$$

This can be derived as follows:

Multiplying (4.2c) vectorial with the vector-product $\mathbf{V} = (\dot{\mathbf{r}} \times \mathbf{r})$ gives

$$\begin{aligned}(V \times r) &= -G\left(m \cdot (M_{\odot} + m)/r^2\right) \cdot [(\dot{r} \times r) \times r] \\ &= -G \cdot \left[(M_{\odot} + m) \cdot \left(-\dot{r}/r + \left(\dot{r}/r^2\right)r\right]\right.\end{aligned} \tag{4.4}$$

The last bracket on the right side is equal to $d/dt(-\mathbf{r}/r)$. Since $d\mathbf{V}/dt = d/dt(\dot{r} \times \mathbf{r}\} = \dot{r} \times \mathbf{r} + \dot{r} \times \dot{r} = 0 \rightarrow V = \text{const.}$ integration of (4.4) yields

$$\mathbf{V} \times \dot{r} = G \cdot (M_{\odot} + m) \cdot r/r + C \tag{4.5}$$

where C is the integration constant.

If this equation is scalar multiplied by $\mathbf{r}$ the left side gives

$$(V \times \dot{r}) \cdot r = V \cdot (\dot{r} \times r) = V^2$$

Equation (4.5) then becomes

$$\begin{aligned}V^2 &= G \cdot (M_{\odot} + m)r + C \cdot r \\ &= G \cdot (M_{\odot} + m)r + C \cdot r \cdot \cos\varphi\end{aligned}$$

where φ is the angle between $\mathbf{C}$ and $\mathbf{r}$.

Solving this equation for r gives

$$r = a/(1 + b \cdot \cos\varphi) \tag{4.6}$$

with $a = V^2/(G \cdot (M_{\odot} + m)$ and $b = C/(G \cdot (M_{\odot} + m))$.

Equation (4.6) describes a conic section in planar polar coordinates $r(t)$ and $\varphi(t)$. It depends on the values of the constants a and b whether this is an ellipse, a parabola or a hyperbola. It turns out that for bound celestial bodies (total energy $E < 0$), the Eq. (4.6) describes ellipses.

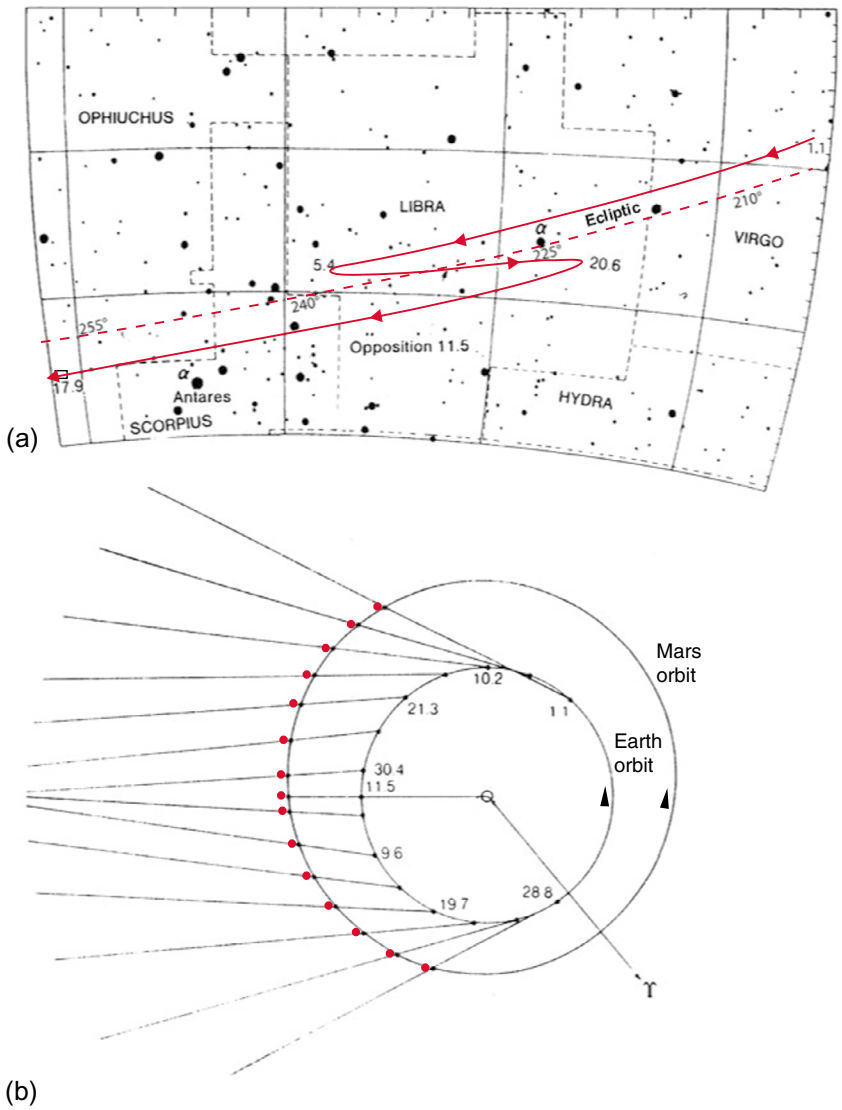


Fig. 4.3 **a** The apparent orbit of Mars in the year 1984, as seen from the Earth. **b** Relative positions of Mars and Earth during the observation time. The projection of the direction Earth → Mars onto the celestial sphere results in the observed loop orbit of Mars (Karttunen et al. Springer Heidelberg)

All planets move around the Sun on elliptical orbits

(First Kepler's law) (Fig. 4.4).

The point with the maximum value of r (largest distance to the Sun) is called the **aphel**, the minimum value of r is the **perihel** (Fig. 4.5).

4.2.2 Second Kepler-Law

The observations give the time-dependent distance $\mathbf{r}(t)$ between Earth and Sun, from which the distances to the center of gravity can be deduced. The radius vector $\mathbf{r}(t)$ can be written as the product of the absolute value r of $\mathbf{r}$ and the unit vector $\hat{\mathbf{e}}$ (Fig. 4.5):

$$\boldsymbol{r}(t) = r \cdot \hat{\boldsymbol{e}}_r(t)$$

The velocity of a planet is then

$$\begin{aligned} \mathbf{v} = \dot{\mathbf{r}} &= \dot{\mathbf{r}} \cdot \widehat{\mathbf{e}}_r + r \cdot \widehat{\mathbf{e}}_r \\ &= \dot{r} \cdot \widehat{\mathbf{e}}_r + r \cdot \dot{\varphi} \cdot \widehat{\mathbf{e}}_\varphi \end{aligned} \tag{4.7}$$

Since the gravitational force is a central force, the orbital angular momentum $\mathbf{L}$ must be conserved (see Vol. 1,

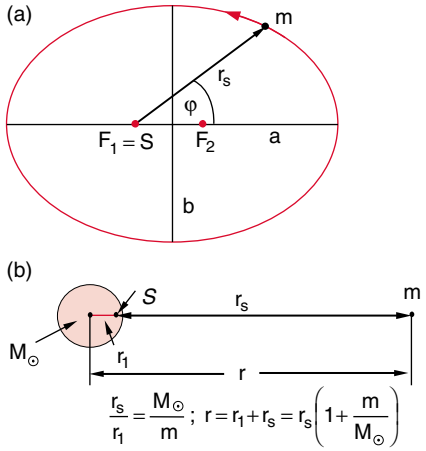


Fig. 4.4 **a** Elliptical orbit of a planet with mass m around the center of mass S, which is located in a focal point of the ellipse. **b** Relation between center of mass distance r_S and relative distance r

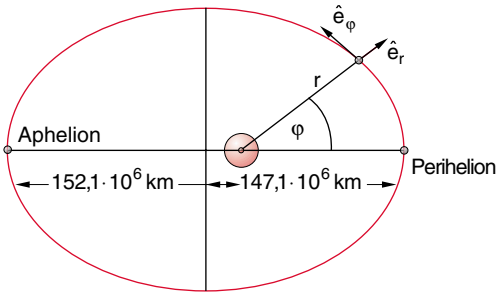


Fig. 4.5 Elliptical orbit of the Earth around the Sun and definition of the unit vectors $\hat{e}_r$ and $\hat{e}_\varphi$

Sect. 2.8. It is perpendicular to the orbital plane (Fig. 4.6a). In polar coordinates, L can be written as

$$\boldsymbol{L} = (\boldsymbol{r} \times m \cdot \mathbf{v}) = m \cdot r^2 \cdot \dot{\varphi} \cdot \hat{\boldsymbol{e}}_z. \tag{4.8}$$

The radius vector r(t) overlaps during the time interval dt the red surface ΔA in Fig. 4.6b, which can be approximated by the triangle $OP_1'P_2'$ with

$$\Delta A = \frac{1}{2} r \cdot \mathbf{v} \cdot \Delta t.$$

In the limit $\Delta t \to 0$ we get

$$\frac{\mathbf{dA}}{\mathbf{dt}} = \frac{\mathbf{1}}{\mathbf{2}} \boldsymbol{rv} = \frac{\mathbf{1}}{\mathbf{2m}} \cdot |\boldsymbol{L}|. \tag{4.9}$$

Since $\mathbf{L} = \mathbf{const.} \rightarrow \mathbf{dA}/\mathbf{dt} = \mathbf{const.}$

This is Kepler's second law:

▶ The radius vector $r(t)$ from the Sun to the planet overlaps in equal time intervals equal areas.

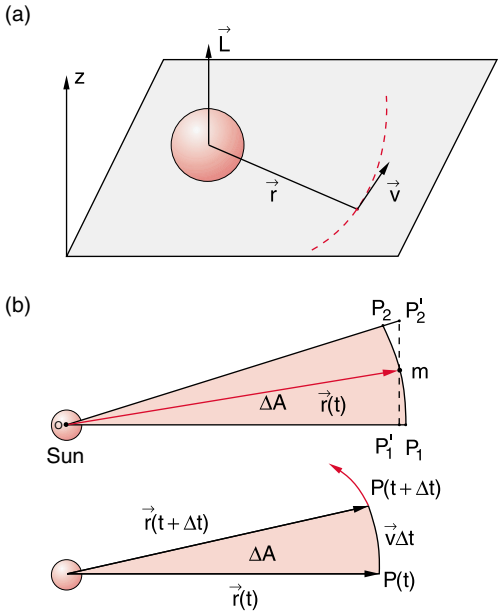


Fig. 4.6 Illustration of Kepler's 2. law: **a** Angular momentum **L** and orbital plane. **b** Kepler's area law

With $|\boldsymbol{L}| \approx m \cdot r \cdot v = \text{const.}$ follows for the velocity of the planet:

$$v \propto 1/r.$$

The planet moves faster in the perihelion ($r = r_{\min}$) than in the aphelion ($r = r_{\max}$).

Integration of (4.9) over a full period gives

$$A = \int \mathrm{d}A = \frac{L}{2m} \int_0^T \mathrm{d}t = \frac{L \cdot T}{2m}. \tag{4.10}$$

The area A of an ellipse with the semi-axes a and b is

$$A = \pi \cdot a \cdot b = \pi a^2 \sqrt{1 - \varepsilon^2} = L \cdot T/2m, \tag{4.11}$$

where $e = \sqrt{(a^2 - b^2)}$ is the eccentricity of the ellipse.

In order to eliminate the angular momentum L, we multiply the equation of motion (4.2c) vectorial with **L** and then scalar with **r**. This gives (see Problem 4.2)

$$L^2 = Gm^2 a \frac{(m + M_\odot)^2}{M_\odot}\left(1 - \varepsilon^2\right). \tag{4.12}$$

Comparing this with the squared Eq. (4.10), we get

$$\frac{T^2}{a^3} = \frac{4\pi^2 \cdot M_\odot}{G \cdot (m + M_\odot)^2}. \tag{4.13}$$

For two different planets with masses m_1, m_2 and major half axes a_1, a_2 of the elliptical orbits, we get the

Third Kepler's Law

$$\frac{T_1^2}{T_2^2} = \frac{a_1^3}{a_2^3}\frac{m_2 + M_\odot}{m_1 + M_\odot} \approx \frac{a_1^3}{a_2^3}. \qquad (4.14)$$

▶ **The ratio of the squared revolution periods T^2 of Planets is equal to the ratio of the cubes of the major axes of their elliptical orbits.**

Example

For our Earth, the numerical values are $T_E = 1$ year, $a_E = 1$ AU, for Mars is $T_M = 1.88$ years, $a_M = 1.52$ AU. $\rightarrow {T_E}^2/{T_M}^2 = 0.283$, ${a_E}^3/{a_M}^3 = 0.283$. ◀

4.2.3 Orbital Elements of the Planets

The orbital planes of the different planets are slightly inclined against the orbital plane of the Earth (see Table 4.2 and Fig. 4.7). The inclination angle i is, however, very small (except for the Mercury orbit). This means that one observes the planets always near to the ecliptic, i.e. the great circle in which the orbital plane of the Earth cuts through the celestial sphere and on which the Sun moves as seen from the Earth.

Table 4.2 Eccentricity ε of the planet orbits, angle γ between the orbital plane and the equator plane of the planets, inclination angle i between ecliptic and orbital plane and sidereal revolution time T^{sid}

Planet	ε	γ	i/°	T/a
Mercury	0.2056	0.01°	7.0	0.24084
Venus	0.0068	−2.6°	3.39	0.61521
Earth	0.0167	23.27°	–	1.0007
Mars	0.0934	24°	1.85	1.88089
Jupiter	0.0485	3.1°	1.31	11.8623
Saturn	0.0556	26.7°	2.49	29.458
Uranus	0.0472	97.9°	0.77	84.01
Neptune	0.0086	28.8°	1.77	164.79

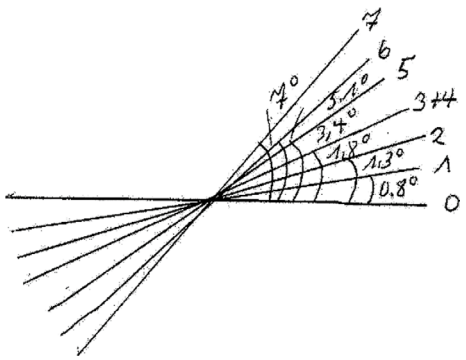


Fig. 4.7 The inclination angles of the planetary orbital planes against the ecliptic. Note that the drawing exaggerates the angles. 1 = Uranus; 2 = Jupiter; 3 = Neptune; 4= Mars; 5= Saturn; 6 = Venus; 7 = Mercury

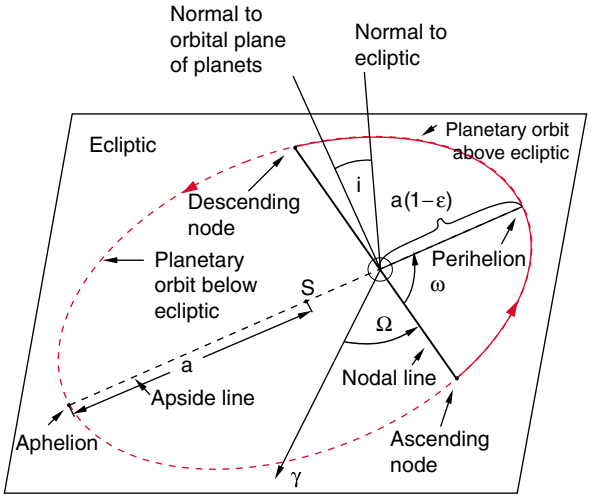


Fig. 4.8 The orbital elements for the description of a planetary orbit. S = center of mass (after Karttunen et al. Springer Heidelberg)

The planets move from a point below the ecliptic from South to North through the ecliptic to a point above the ecliptic. The intersection with the ecliptic is the *ascending node*, (Fig. 4.8). The opposite point where the planet intersects the ecliptic from north to south is the *descending node.* The line through the two nodes is the intersection line between the planes of the planet orbit and the Earth orbit (ecliptic).

The orbit of a planet is determined by six characteristic quantities:

It would be obvious to choose the radius vector $\boldsymbol{r}(t)$ and the velocity $\boldsymbol{v}(t)$ of the planet at a time t. For observations, it is more convenient to use the following characteristic observational quantities:

- The major semi-axis of its elliptical path,
- The eccentricity ε,
- The inclination of its orbital plane against the ecliptic,
- The angle Ω between vernal equinox and ascending node,
- The angle ω between ascending node and perihelion of the elliptic orbit,
- The perihelion time, i.e. the time when the planet has reached the perihelion.

One distinguishes between

(a) *lower planets* (Mercury, Venus with a < 1 AU) and
(b) *upper planets* (Mars, Jupiter, Saturn Neptune and Uranus with a > 1 AU) (Table 4.3).

If a planet is located on the line of sight Earth-Sun at the opposite side of the Sun as seen from the Earth, its location

Table 4.3 Classification into inner (a < 1 AU) and outer planets (a > 1 AU)

Classification solid or gaseous	Planets and their symbols		Classification distance to Sun
Earth-like solid	Mercury	☿	Lower planets $a_P < a_E$
	Venus	♀	
	Earth	♁	
	Mars	♂	
Jupiter-like Gaseous	Jupiter	♃	Upper planets $a_P > a_E$
	Saturn	♄	
	Uranus	♅	
	Neptune	♆	

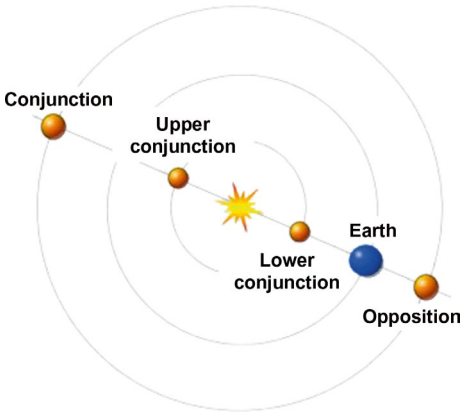


Fig. 4.9 Constellation of opposition and conjunction of an outer planet and the two conjunction positions of an inner planet as seen from the Earth

is called **opposition**, if it is located on the same side as the Sun, it is in **conjunction**. As is illustrated in Fig. 4.9, the lower planets are never in opposition, but they have two conjunction points, the lower and upper conjunction, while for upper planets there is one conjunction point and one opposition point (Fig. 4.9). At the lower conjunction point,

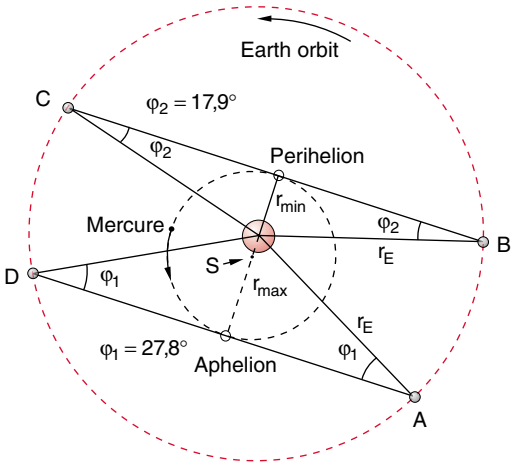


Fig. 4.10 Maximum elongation angle φ (angle Sun-Earth-planet) for the example of Mercury

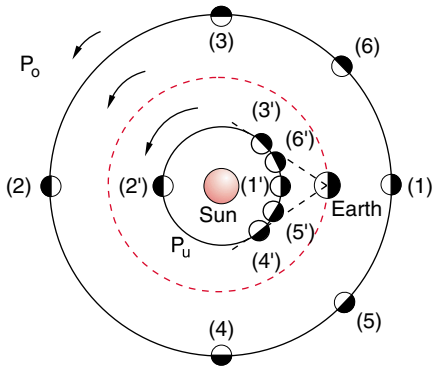


Fig. 4.11 Illumination phases for an outer and for an inner planet

the Venus has its smallest distance to Earth. During their conjunction, the planets are not visible because they either reach this point at daytime and are therefore outshined by the Sun (upper conjunction) or they show us the dark side (lower conjunction).

The lower planets can be seen from the Earth only in an angular range $\varphi < \varphi_{\max}$ against the line of sight Earth-Sun. During a year, the angle φ can have several maxima, which occur when the line of sight Earth-planet becomes the tangent to the planet orbit (Fig. 4.11).

Example

The two conjunction points of a lower planet are for the eccentric orbit of Mercury ($\varepsilon = 0.2$, $a = 0.387$ AU). ◀

$$\begin{aligned} \sin\varphi_1^{\max} &= \frac{r_{\max}}{r_E} = \frac{a(1+\varepsilon)}{r_E} \approx \frac{0.47\,\text{AU}}{1\,\text{AU}} = 0.47 \\ &\Rightarrow \varphi_1 = 28°, \\ \sin\varphi_2^{\max} &= \frac{a(1-\varepsilon)}{r_E} = 0.31 \\ &\Rightarrow \varphi_2 = 18°. \end{aligned} \tag{4.15}$$

Mercury, therefore, can be only seen shortly after sunset or shortly before sunrise.

For the observation points A and C, Mercury appears on the evening sky, shortly after sunset, for the points B and D, it appears on the morning sky shortly before sunrise.

A similar situation occurs for Venus. Before it reaches the lower conjunction, it can be seen on the evening sky. After having passed the lower conjunction, it appears on the morning sky. Therefore, Venus, which is easily visible as a very bright object, is called *Evening- and Morning Star* (*Lucifer*).

The observer on Earth sees only that illuminated part of the planet that faces the Earth. Therefore, one observes periodic planet phases (Fig. 4.12) as for instance the Venus phases depicted in Fig. 4.13. This figure also illustrates the different sizes of the Venus disc (at the same telescope

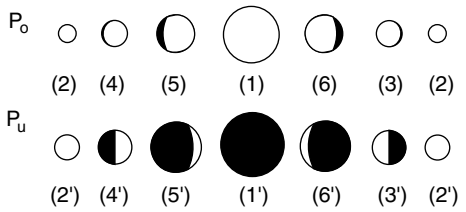


Fig. 4.12 Illumination phases P_0 for an outer planet and P_i for an inner planet for the positions given in x Fig. 4.11 (after Gondolatsch et al. Astronomie (Klett Stuttgart 1995)

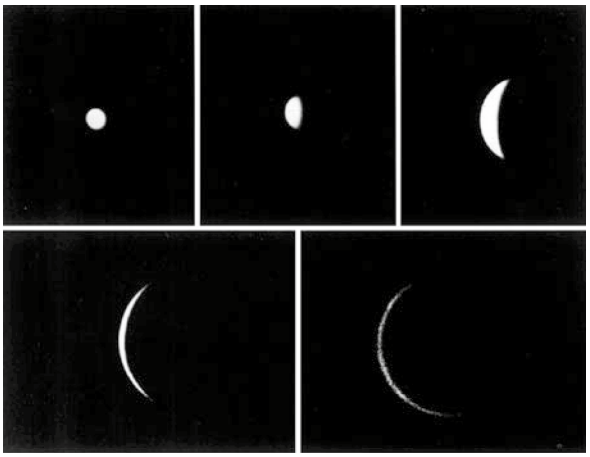


Fig. 4.13 The planet Venus at different distances, observed from the Earth at the positions 2; 3; 6 and between 6 and 1 and in 1 (from Gondolatsch Astronomie Klett Stuttgart 1995)

magnification) due to the different distances Earth-Venus. The apparent diameter of the Venus image changes with its phase. It is minimum when Venus appears as a full disc because the distance is maximum at this phase.

Measurements of the Venus phases allow the determination of the distance Earth-Venus (Fig. 4.14). The phase $P = b/2R$ is defined as the illuminated fraction of the Venus equator, which would appear to us as the diameter 2R, if we see the full disc of Venus. From Fig. 4.14, we can deduce

$$\frac{r}{r_E} = \frac{\sin(\alpha+\varphi)}{\sin\varphi} \Rightarrow r = 1\text{AU} \cdot \frac{\sin(\alpha+\varphi)}{\sin\varphi} \tag{4.16}$$

The angle α is given by the position of Venus and is the angle between the lines Earth-Sun and Venus-Sun. The angle φ is, according to Fig. 4.13,

$$\sin(\varphi - 90°) = \frac{R-b}{R} = 1 - 2P \tag{4.17}$$

where R is the radius of Venus.

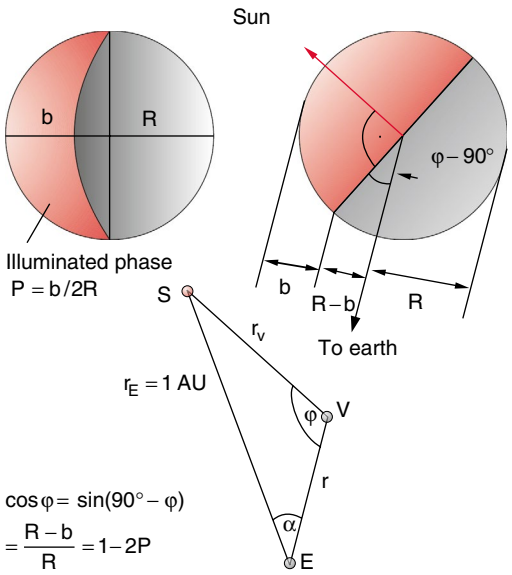


Fig. 4.14 Determination of the distance Earth-Venus from the Venus phases

4.2.4 The Periods of the Planets

According to the 3rd Law of Kepler the circulation periods T of the planets increase with their distance r from the Sun as

$$T \propto a^{3/2}$$

where the distance r is expressed by the major semi-axis of the elliptical orbit. In Table 4.2, the periods of all planets are compiled. For instance, the period of Jupiter is $T = 11.86$ years.

The average velocities of the planets are given by

$$v_i = \frac{\text{circumference}}{\text{circulation period}} \approx \frac{2\pi a_i}{T_i} \propto \frac{1}{\sqrt{a_i}} \tag{4.18}$$

they decrease as $v_i \propto a^{-1/2}$.

Example

The average velocity of the Earth is

$$v_E = \frac{2\pi}{1a} \cdot 1\,\text{AU} \approx 29.8\,\text{km/s}\,.$$

The average velocity of Jupiter is $v_J = \left(\frac{2\pi}{11.9y}\right) \cdot 5.2\,\text{AU} = 13.1\,\text{km/s}$. ◀

One must distinguish between the **sidereal period** T_{sid} where the planet performs a full circle around the Sun and

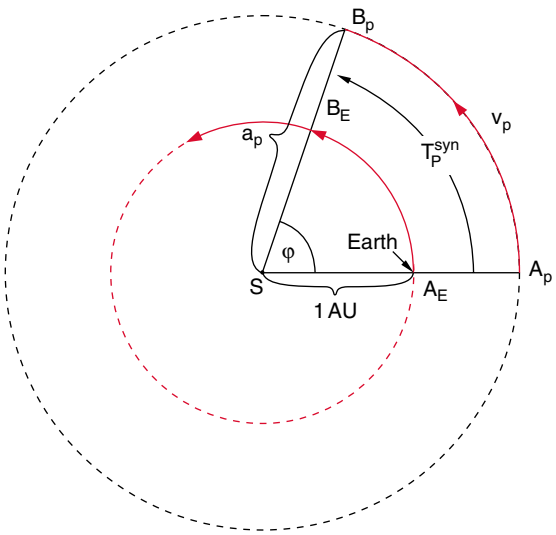


Fig. 4.15 Illustration of the relation between sidereal and synodal revolution period

the **synodal period T_{syn}**, which is the time between two successive conjunctions (or oppositions) for an observer on the moving Earth (Fig. 4.15).

We will illustrate this difference in Fig. 4.15 for an upper planet. The mean angular velocities of Earth ω_E and planet ω_P are

$$\omega_E = \frac{2\pi}{T_E^{sid}}; \omega_P = \frac{2\pi}{T_P^{sid}}. \tag{4.19}$$

The planet should be at time $t=0$ in opposition to the Sun and the Earth is then located at the point A_E on the line Sun-planet. After the synodic period T_P^{syn} at the next opposition Sun-Earth and planet are again on the line Sun-Earth-planet. Since Earth and planet both move in the same direction but the velocity of the Earth is larger than that of the outer planet, the second opposition occurs after the time

$$T_{syn} = T_{sid} + \varphi \cdot a_E/v_E = \varphi \cdot a_P/v_P \tag{4.20}$$

With $v_E = 2\pi \cdot 1AU/1y$ and $v_P = 2\pi \cdot a_P/T_P^{sid}$, we get for an upper planet

$$\frac{1}{T_P^{syn}} = \frac{1}{T_E^{sid}} - \frac{1}{T_P^{sid}}. \tag{4.21a}$$

For a lower planet, which moves faster than the Earth, we get

$$\frac{1}{T_P^{syn}} = \frac{1}{T_P^{sid}} - \frac{1}{T_E^{sid}}. \tag{4.21b}$$

Note: **Astronomical observations give the synodic period. It has to be converted into the sidereal period, which is used in Keplers laws.**

Comment: The synodic orbital frequency $\nu_P^{syn} = 1/T_P^{syn}$ of a planet corresponds to the difference frequency $\nu_E^{sid} - \nu_P^{sid}$ for an upper planet and to $\nu_P^{sid} - \nu_E^{sid}$ for a lower planet. One can therefore regard the orbital frequencies of the planets and their overtones as a celestial harmony, as was proposed by Kepler, who titled his book about the motion of the planets as "*harmonices Mundi*" (the harmony of the world).

4.2.5 Size, Mass and Mean Density of the Planets

When observing the planets through a telescope they appear as small discs with an angular diameter $\Delta\alpha$ which gives the true diameter

$$D = r \cdot \Delta\alpha$$

if their distance r from the Earth is known. The distance r between Earth and planet changes continuously because the velocities of the planets differ. For the lower planets Mercury and Venus, r is minimum for the lower conjugation, where the planet, however, is not visible (Fig. 4.10).

With the angle φ between the lines Earth-planet and Sun-Earth (Fig. 4.16) the cosine rule for the triangle ESP gives the relation

$$\begin{aligned} &\boldsymbol{r}_E + \boldsymbol{r}_{EP} = \boldsymbol{r}_P \\ &\Rightarrow r_E^2 + r_{EP}^2 + 2r_E r_{EP}\cos\varphi = r_P^2, \end{aligned} \tag{4.22}$$

with the known distances $r_E = 1$ AU and r_P (from the measured period of P) and the measured angle φ one obtains the distance r_{EP} between Earth and planet at the time of the measurement.

For the neighboring planets *Venus* and *Mars,* more accurate distance determinations are possible with time of flight measurements of electromagnetic pulses (radar or laser pulses) from Earth to planet and back (see Sect. 3.x).

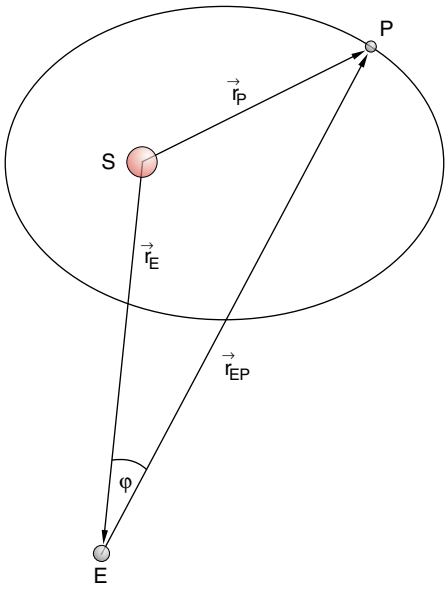


Fig. 4.16 Measurement of the distance of a planet from the Earth

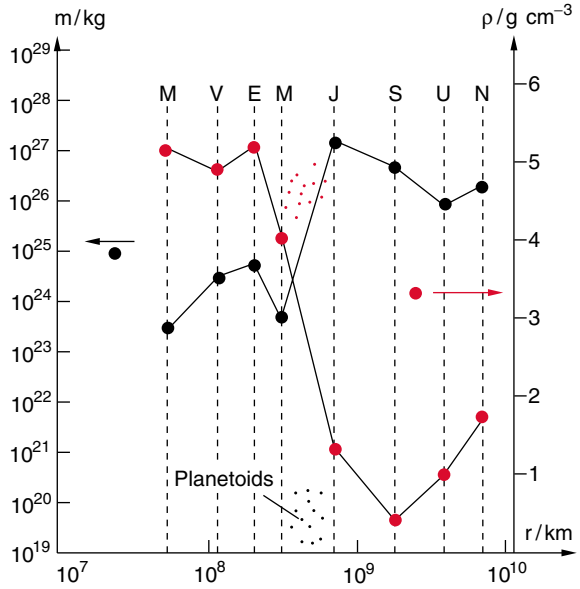


Fig. 4.17 Masses (black points) mean densities (red points) and distances from the Sun (abscissa) for the planets

Measurements of the planet diameters parallel and perpendicular to the rotation axis give the deviation of the planet from a spherical form.

The measurements prove that Mercury and Venus are essentially of spherical size while Earth and Mars are rotational ellipsoids with small oblateness (see Table 4.1).

The mass of a planet with at least one moon can be directly obtained from the distance of the moon from the planet and its period on the orbit around its planet, using Kepler's third law.

Example

The Earth moon has a period of $T = 27.32166$ d. The mean distance from the Earth is $r = 384400$ km. From ◀

$$\begin{aligned} \frac{mv^2}{r} &= G \cdot \frac{m \cdot M_\mathrm{E}}{r^2} \\ \Rightarrow M_\mathrm{E} &= \frac{r \cdot v^2}{G} = \frac{4\pi^2}{G} \cdot \frac{r^3}{T^2} \\ \Rightarrow \boldsymbol{M}_\mathrm{E} &= 6 \cdot 10^{24}\mathbf{kg}. \end{aligned} \tag{4.23}$$

For Mercury and Venus, which have no moons, the mass had been determined in former times from the gravitational force on the neighboring planet, which cause measurable perturbations of its orbit.

Much more accurate determinations of their mass could be nowadays achieved from the deviation of satellite orbits passing close by or from the period of satellites circling around the planet.

The masses of the planets and planetoids are plotted in Fig. 4.17 on a logarithmic scale, and their densities on a linear scale. The figure illustrates that the inner planets form a group with a mean mass of about 10^{24} kg, while the outer planets form a group with a much larger mass around 10^{27} kg. Below both groups, the planetoids are arranged with masses around 10^{20} kg.

From the radius R of a planet and its mass m the mass density ρ can be calculated as

$$\overline{\varrho} = \frac{m}{\frac{4}{3}\pi R^3} \tag{4.24}$$

Figure 4.17 shows that the inner planets have mean densities around $\overline{\varrho} \approx 4{-}5\,\mathrm{kg/dm^3}$, whereas the densities of the outer planets are with $\overline{\varrho} \approx 1.4\,\mathrm{kg/dm^3}$ smaller by a factor 3–4 and are comparable with the mean density of the Sun.

▶ The inner planets have small masses but high densities, while the outer planets have large masses but small densities.

The inner planets are called *Earth-like*, the outer planets *Jupiter-like*.

4.2.6 The Names of the Planets

The names of the planets originate from the Roman pantheon. The largest planet *Jupiter* got the name of the chief of all gods (the Greek name is *Zeus*), his father was

Saturn (Greek name Cronos), designated by the symbol of a sickle, because he killed his father *Uranus* (the grandfather of *Jupiter*) with a sickle. *Neptune* is the god of the sea, *Mars* the god of war. *Mercury* is the messenger of the gods. *Venus,* as the bright evening and morning star got the name of the Roman goddess of love.

4.2.7 Energy Budget of the Planets

Planets are not self-luminous bodies. They reflect and scatter the light they receive from the Sun. This occurs in their atmosphere and for the inner planets also at the solid surface.

The ratio A of diffuse backscattered and reflected radiation power to the vertical incident radiation power

$$A = \frac{S_{\text{back}}}{S_{\text{in}}} \tag{4.25}$$

is called the **albedo** of the planet (from the Latin albus = white).

In Table 4.4, the values of A are listed for the different planets and some moons. For comparison, the albedo for fresh snow and for clouds are also given.

A planet with the distance r from the Sun receives the radiation power density (=intensity)

$$S(r) = SC \cdot \left(\frac{1\,\text{AU}}{r}\right)^2 = \frac{R_\odot^2}{r^2} \cdot \sigma \cdot T_\odot^4, \tag{4.26a}$$

where $SC = 1.37\,\text{kW/m}^2$ is the solar constant, i.e. the radiation power that falls onto 1 m^2 outside the Earth atmosphere, and σ is the Stefan-Boltzmann constant.

The total radiation power emitted by the Sun is $4\pi \cdot R_\odot^2 \cdot \sigma \cdot T_\odot^4$ where $T_\odot$ is the temperature of the Sun surface.

Table 4.4 Albedo for the planets of the solar system, the Earth moon and the Jupiter moon Enceladus

Planetary	Albedo
Mercury	0.09
Venus	0.76
Earth	0.31
Mars	0.25
Jupiter	0.50
Saturn	0.34
Uranus	0.30
Neptune	0.29
Earth moon	0.11
Enceladus	0.8
For comparison	
Fresh snow	0.8–0.9
Clouds	0.6–0.9

Note that the total radiation power received by the Earth is $\frac{\text{dW}}{\text{d}t} = SC \cdot \pi R^2$ [W]. The radiation intensity averaged over the whole surface is then

$$S(r) = SC \cdot \frac{\pi R^2}{4\pi R^2} = \frac{1}{4} \cdot SC.$$

Furthermore, for planets with an atmosphere, part of the incident radiation is reflected and backscattered by clouds and aerosols in the atmosphere.

At a mean albedo A, the radiation power absorbed by the planet with radius R and its atmosphere is

$$\frac{\text{d}W_1}{\text{d}t} = \pi R^2 (1 - A) \cdot S(r) \tag{4.26b}$$

Furthermore, a planet can possess inner energy sources, which are for the Earth for instance the decay of radioactive substances in the inner part of the Earth. For the gaseous planet part of their inner energy comes from gas contraction and condensation. This energy can be transported to the surface by convection.

With the heat flux Q_i [W/m^2] from the inwards to the surface the total energy transport is

$$\frac{\text{d}W_2}{\text{d}t} = 4\pi R^2 Q_i. \tag{4.27}$$

If the rotation of the planet is fast enough, the surface of the planet has everywhere the same temperature T_p and the planet radiates, according to the Stefan-Boltzmann radiation law, the energy

$$\frac{\text{d}W_3}{\text{d}t} = 4\pi R^2 \sigma \cdot T^4 \tag{4.28}$$

per sec into space.

Equilibrium is reached when the absorbed power is equal to the emitted power. This gives the equation

$$\frac{\text{d}W_3}{\text{d}t} = \frac{\text{d}W_1}{\text{d}t} + \frac{\text{d}W_2}{\text{d}t}$$

$$4\pi R^2 \cdot \sigma \cdot T_\text{P}^4 = \pi R^2 (1 - A) \cdot S(r) + 4\pi R^2 \cdot Q_i \tag{4.29}$$

For the Earth the last term is negligible, and we therefore obtain the planet's temperature when inserting S(r) from (4.26a)

$$T_p = \left[(1 - A) \cdot SC \cdot 1\,\text{AU}/4r^2 \cdot \sigma\right]^{1/4} \tag{4.30}$$

Example

For the rotating Earth without its atmosphere, we get with A = 0.31, SC = 1361 W/m^2, r = 1 AU, $\sigma = 5.670 \cdot 10^{-8}\,\text{SW m}^{-2}\,\text{K}^{-4}$

T = 253.7 K = −19.5 °C. ◀

Without atmosphere it would be therefore very cold on Earth.

Taking into account the small energy flux from the inner part of the Earth to the surface (3 W/m^2), the temperature would increase to T = −18 °C, still below the freezing point of water. It was therefore no life on Earth possible because there would be no liquid water.

For a non-rotating planet, there would be a stationary temperature T_{eff} on the dayside which is determined by the equilibrium between incident radiation and emitted radiation and eventual heat conduction to the nightside.

Without atmosphere, the nightside will cool down so much, that we can neglect the emission of radiation from this side. We then obtain for the dayside from the equilibrium condition

$$2\pi R^2\sigma \cdot T_p^4 = \pi R^2(1-A)S(r) + 2\pi R^2 Q_i \quad (4.31)$$

Neglecting the heat conduction term Q_i (which amounts only to 3 W/m^2 for the Earth), the stationary temperature of a nonrotating planet without atmosphere becomes

$$T_p^{day} = (1\text{-}A)^{1/4}(R_\odot/r)^{1/2}\cdot T_\odot \quad (4.32)$$

where the Sun radius $R_\odot$ and the distance r are given in *m*.

On the nightside of the nonrotating planet, the equilibrium between radiation energy radiated into space and heat conduction Q_i pro m^2 from the dayside and from the interior is

$$2\pi R^2 \cdot \sigma \cdot T_P^4 = Q_i \cdot \pi R^2$$

The temperature becomes

$$T_{eff}^{night} = \sqrt[4]{Q_i/2\sigma}. \quad (4.33)$$

Example

For a nonrotating Earth without atmosphere, the equilibrium temperatures are with $r = 1$ AU, A = 0.3, SC = 1.37 kW/m^2, neglecting Q_i

$$T_{eff}^{day} = 351\ \text{K} \Rightarrow T_0 = -22\,^\circ\text{C}$$

$$T_{eff}^{night} = 40\ \text{K} \Rightarrow T_0 = -233\,^\circ\text{C}$$

The day temperature is slightly lower than for a rotating Earth. ◀

For a fast rotating planet, the temperature is nearly the same all over the planet's surface and it therefore emits the same radiation power $\left(\frac{dS}{4\pi R^2}\right)\cdot\sigma\cdot T^4$ from each surface element dS. The effective temperature is now smaller by a factor $\sqrt{2}$ because the incident radiation falls in the average onto the planets cross-section πR^2 but the emitted radiation is proportional to the surface $4\pi R^2$. Therefore, the temperature should be smaller by the factor $\sqrt[4]{4} = \sqrt{2}$.

The measured temperatures are, however, higher than these calculated values. The reason for this discrepancy is the atmosphere of the planets, which absorbs the infrared radiation emitted from the surface of the planet and is therefore heated up. Since the main constituents of the atmosphere, namely, nitrogen N_2 and oxygen O_2, do not absorb in the infrared, the minor concentration of trace molecules such as CO_2, NH_4 and H_2O are responsible for the absorption. Small changes of their concentration will affect the total absorption by the atmosphere and will therefore also change the temperature on the surface of the Earth.

Due to this absorption of the infrared radiation emitted from the surface of the planet, the first term on the left side of (4.29) will become much smaller and the temperature of the planet increases.

Example

For the Earth with its atmosphere, the mean temperature would rise to T = +15°. ◀

Planets without or with a thin atmosphere show strong day and night temperature changes, while planets with a dense atmosphere have only small differences $T_{day} - T_{night}$.

In Fig. 4.18, the incident and emitted radiation power and the energy balance of the Earth are illustrated with the different contributions, due to the reflection and absorption by the Earth surface and by clouds. This figure shows on the left side, which part of the incident radiation (100%) is absorbed and scattered in the atmosphere. The absorption is due to Ozone O_3 in the higher atmosphere (3%) by water, dust and aerosols in the troposphere (17%) and by the Earth surface (22%). The total absorption of the incident Sun radiation, therefore, amounts to 42%, while 28% are reflected and 30% are scattered.

On the right side of Fig. 4.18, the different sinks of the radiation emitted by the Earth surface are depicted. This illustrates that the major part is absorbed by trace molecules in the atmosphere or are backscattered to the Earth surface. These trace molecules are therefore essential for the energy balance between incident and emitted radiation and the Earth temperature depends critically on their concentration.

Note: The real temperature on Earth is not the same all over the surface. This is partly due to the inclination of the Earth axis, which causes that the temperature is in the north summer higher on the northern hemisphere and in the north winter higher on the southern hemisphere. Local pressure changes of the atmosphere cause winds which transport energy and affect the local temperature and can cause quite large temperature differences for different locations.

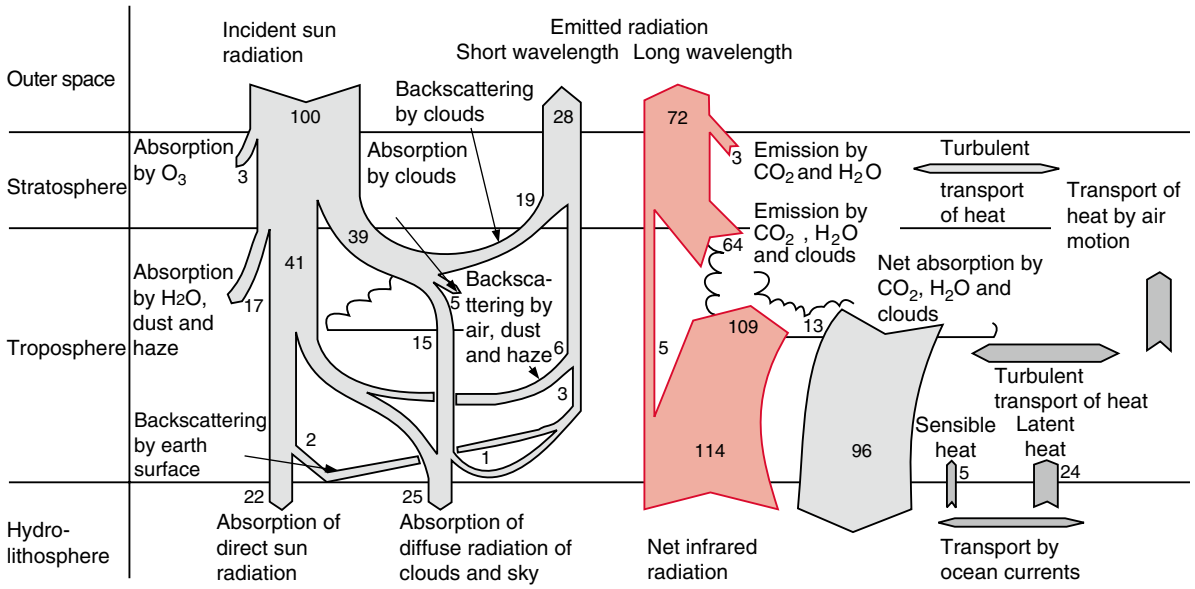


Fig. 4.18 Energy balance of the Earth and its atmosphere (after S. H. Schneider R. Lander: The Coevolution of Climate and Life. Sierra Club books San Francisco 1989) [4]

4.3 The Inner Planets and Their Moons

The four inner planets *Mercury, Venus, Earth and Mars* all consist of solid material. While Mercury has essentially no atmosphere $(p = 10^{-15}$ bar$)$, *Venus* has a very dense atmosphere ($p \geq 90$ bar) compared to the atmosphere of the Earth (p = 1 bar). Mars, on the other side has a much lower density (p = 7 mbar). Mercury and Venus have no moon, the Earth has one and Mars has two moons (*Deimos* and *Phobos*).

The surfaces of the inner planets which have been meanwhile accurately photographed are for all inner planets similar. All of them show crater areas and volcanic basins. The craters had been formed by meteoroid impacts while the volcanic basins were formed by melting processes in the inner part of the planet where the solid crust had been broken and the liquid lava passing out, solidified and formed the plane basins. The crater areas prove that the crust of the planet must be thick enough to withstand the impact of meteorites. Examples are Mercury and the Earth moon, which show quite similar surfaces (Fig. 4.19).

4.3.1 Mercury

The orbit of Mercury has the largest eccentricity $\varepsilon = 0.2056$ of all planets. Due to the gravitational perturbation by the next planet Venus, the total force acting on Mercury is no longer a pure central force. This has the effect that the major axis of the elliptical orbit is not space-fixed but precesses in 100 years by the angle 531″ (perihelion precession, Fig. 4.19).

The observed perihelion precession amounts, however, to 574″/100y. The difference of 43″/100y could be

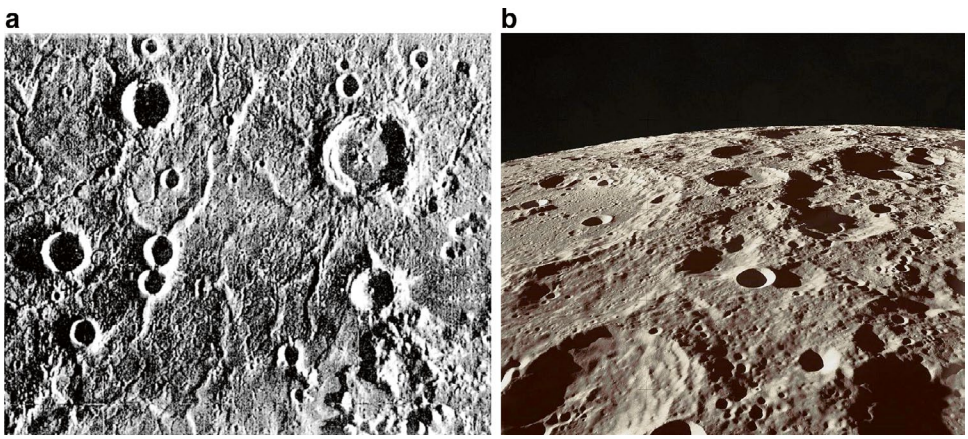


Fig. 4.19 **a** Details of the surface of Mercury. Inner part of the Caloris basin, photographed by the space probe Mariner 10. The diameter of the big crater is 60 km (NASA). **b** Part of the moon surface (NASA)

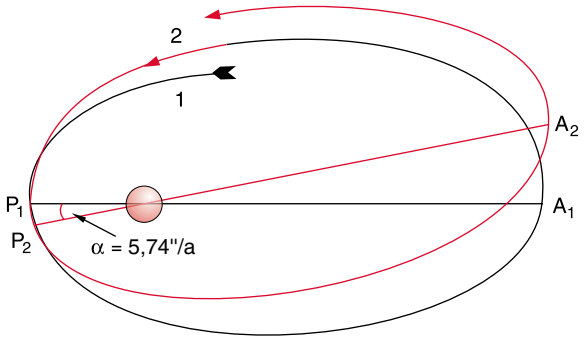


Fig. 4.20 Perihelion precession of the Mercury orbit

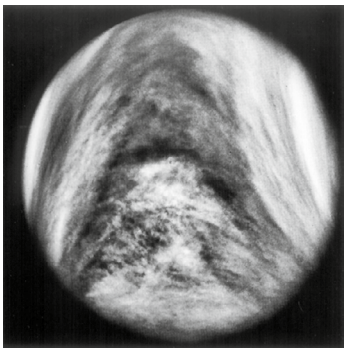


Fig. 4.21 Photo of Venus with its dense clouds (NASA)

explained only 1905 by Einstein with the general relativity theory. This was the first convincing experimental proof of the validity of this theory (Fig. 4.20).

Mercury is with its diameter $D = 4878$ km the smallest planet and only slightly larger than the Earth moon. Its mean density is $\overline{\varrho} = 5.44\,\mathrm{kg/dm^3}$, which induces the assumption that Mercury has an iron kernel with a diameter of about 1800 km. Radar measurements with radio telescopes gave the information that mercury rotates with a period of von $T_{\mathrm{sid}} = 58.65$ d. Its synodal rotational period which determines the length of day and night, is $T_{\mathrm{syn}} = 176$ d. Since its revolution time around the Sun is 88d, *one day on mercury lasts two mercury years!*

The ratio of sidereal rotation period to revolution period is 2:3. This 2:3-resonance is probably due to the deceleration of an initially faster rotation, caused by the friction of tides, induced by the Sun. The deceleration can be effective because Mercury has a small hump of about 10^{-5} R_{mercury}. Because of this deviation from a sphere, the gravitational force between Sun and Mercury causes a torque, which is decelerating the rotation of Mercury.

The USA-space probe Mariner has sent detailed photographs of the Mercury surface (Fig. 4.19a). Besides many meteorite impact craters, there are large volcanic basins, where the largest one is the *Caloris basin* with a diameter of 1300 km. It has been probably formed by the impact of a very massive body which caused the fracture of the crust with subsequent flooding of the basin with liquid lava. Vulcanism must have played a much greater role on Mercury than on the Earth moon as can be concluded from the many volcanic structures such as fractures and breakings of the crust, lava-seas, etc.

Since Mercury has no atmosphere which can store the heat energy very large temperature differences occur during the long day- and night times. The temperature at noon rises to 467 °C and sinks during night time to −183 °C, as has been measured by the space probe Mariner.

4.3.2 Venus

Venus is, after Sun and Moon, the third brightest object on the sky. This is not only due to the fact that Venus has from all planets the smallest distance to Earth during its nearest approach and therefore scatters more light onto the Earth but also because of its high reflectivity (albedo $A = 0.7$) which is caused by its dense atmosphere with highly reflecting clouds (Fig. 4.21). The dense atmosphere causes a pressure of 90 bars at the Venus surface. It mainly consists of CO_2 (96%), nitrogen N_2 (3.5%) with the rest of 0.5% distributed among sulfur dioxide SO_2, sulfur acid H_2SO_4, water H_2O and argon. The Venus surface is always covered with dense clouds which makes it impossible to view its solid surface from outside. Only about 2% of the incident sunlight reaches the solid surface [9].

Russian and American space probes (*Venera* and *Pioneer*) could measure the pressure and temperature progression in the Venus atmosphere. Some probes equipped with parachutes could land on the solid surface and have sent detailed pictures of the structure and morphology of the surface (Fig. 4.22). The convey besides some mountains, which rise up to 10 km, a surprisingly flat structure. About 60% of the surface are plains. There are some huge craters with diameters up to 270 km (Fig. 4.23), which must have been formed before the Venus atmosphere had reached its present density.

At the solid surface, the atmospheric pressure is 90 bars and a temperature of 730 K has been measured. This high temperature is caused by the dense CO_2 atmosphere, which absorbs the infrared radiation emitted by the solid surface (greenhouse effect). The dense atmosphere prevents the energy emission from the surface. Therefore, the temperature differences between day and night are small.

The cloud density increases with increasing height and reaches its maximum at about 50 km altitude.

With RADAR measurements from the Earth Doppler-shifts of the signals reflected from the edges of Venus

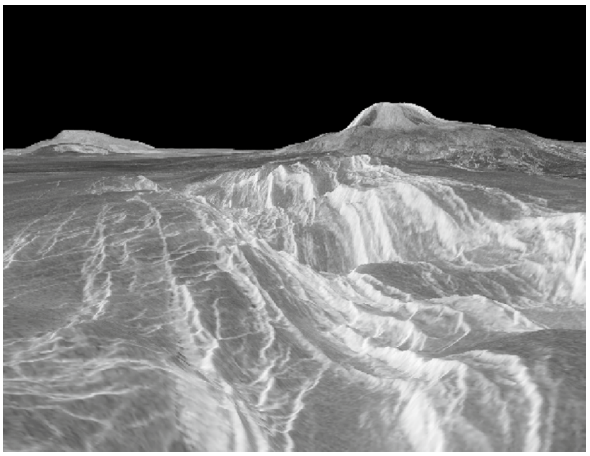


Fig. 4.22 Rift valley on the Venus surface. The picture was computer-composed from photos of the Russian Venera probe, and radar pictures of the Magellan Mission. The height scale is exaggerated in order to illustrate the folding of the Venus crust (Magellan Mission NASA)

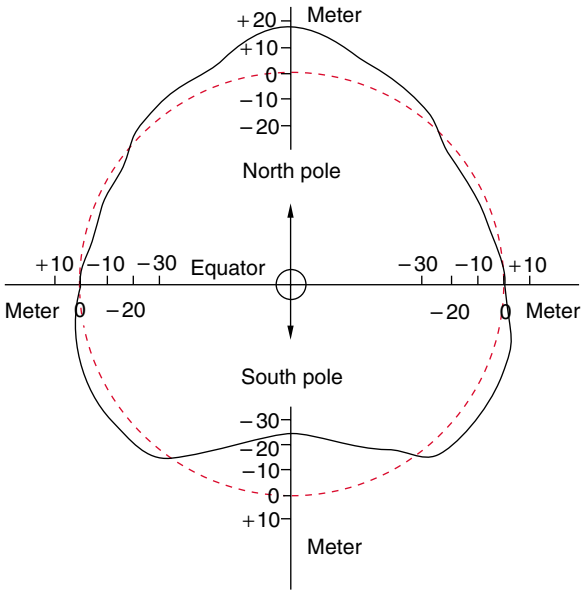


Fig. 4.24 Deviation of the Geoid (solid line) from the rotational ellipsoid (dashed line) with the pole oblateness $(a - b)/a = 1/298.25$. The deviations are exaggerated by a factor 80.000

proved that Venus rotates retrograde, i.e. the spin angular momentum points into the opposite direction as the orbital angular momentum. The rotation period is 243.1 d.

▶ The length of a Venus day (synodal rotation period) is with 116.75 d shorter than the sidereal rotation period of 243.1 d, due to the retrograde rotation.

Venus has no magnetic field, which is probably due to its slow rotation, which is not fast enough to generate a magnetic field by the dynamo effect (see Vol. 2, Sect. 3.6).

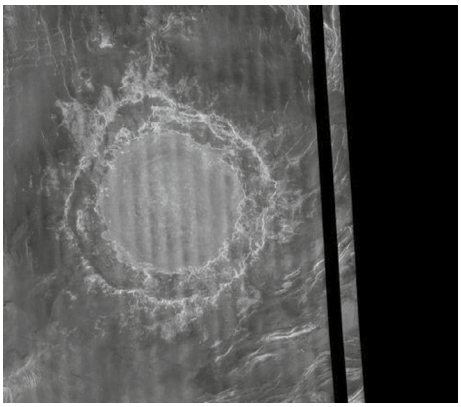


Fig. 4.23 The Venus crater *Mead* (NASA)

4.3.3 The Earth

Our Earth is, of course, from all planets the one which has been most detailed explored. Its outer form can be approximately described as a distorted rotational ellipsoid with an equator radius of $r_{equ} = a = 6378\,km$ and a pol radius, (which equals half of the distance between the poles) of $r_{pol} = 6357\,km$. The ellipticity $(a - b)/a = 0.0033$ is about 0.3%. More detailed measurements have proved that the real form of the Earth is a **geoid**, which is shown in exaggerated form not to scale in Fig. 4.24. The numbers give the real deviations from the rotational ellipsoid in meters.

4.3.3.1 The Inner Structure of the Earth

The inner structure of the Earth can be described by a shell model (Fig. 4.25). The central part (inner core) has a radius of about 1220 km and consists of iron and nickel. The next layer is the outer core, which is composed of liquid metals. The mean density of the core is $\overline{\varrho} = 13–14\ kg/dm^3$. It is higher than the density of iron at normal pressure and can be explained by the compression caused by the extremely high pressure of $p = 3 \times 10^{11}\,Pa = 3 \times 10^7$ bar in the core.

Above the core is the Earth mantle, which consists of hot solid rocks (iron silicates) and due to its high temperature and pressure it forms a viscous material. Its radial range is 2900 km and contains about 67% of the total mass and 87% of the total volume of the Earth. At mid-ocean ridges, partial melting of the mantle can occur which produces the ocean crust, while at subduction zones the continental crust is formed.

The outer zone is the Earth crust (lithosphere), which is with 35 km relatively thin and its density is lower than that of the mantle. It therefore "floats" on the viscous mantle and forms large plates. One distinguishes between the continental crust, which contains mainly silicon, oxygen and magnesium and the much thinner oceanic crust which is mainly composed of silicon, oxygen and aluminum.

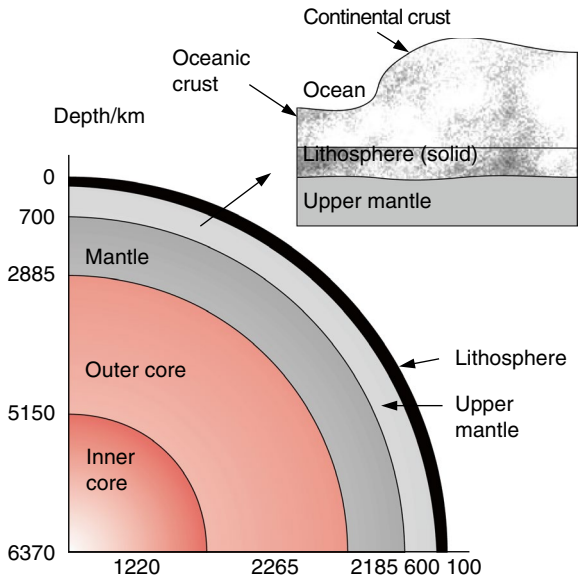


Fig. 4.25 Schematic picture of the Earth structure (the numbers on the abscissa give the thickness of the different layers in km)

The lithosphere is not a unique cohesive layer but is fractured into seven large continental plates and some smaller plates (Fig. 4.26), which slowly move, where the heavier plate slides below the lighter neighboring plate. This causes earthquakes with a strength that depends on the magnitude of the plate shift.

The movement of the plates is driven by magma currents below the plates, which lifts the upper plate and forms ranges of mountains. For instance, the Himalaya mountains have been formed 50 million years ago by collisions between the continental Indian plate and the Eurasian plate.

The question is now, how can we find out about the interior of the Earth?

The up to now deepest bore is only 12 km deep and the analysis of the drill core gives only information about the outer part of the lithosphere. Therefore, other methods have to be applied. One of them is the analysis of earthquakes, which originate at the epicenter but can be detected at various locations on Earth because it creates seismic waves (acoustic waves), which spread out into all directions (Fig. 4.27). If the velocity and the damping of these acoustic waves can be measured, much information about the material on the way between source and detector can be obtained.

The acoustic waves are reflected and refracted at the boundary layers between two different shells. Their behavior depends on the polarization of the waves. We distinguish between transverse and longitudinal acoustic waves. Since the shear modulus of liquids is zero, transverse waves cannot penetrate into the liquid zone of the Earth interior. They are reflected at the boundary (Fig. 4.27a) Therefore, only longitudinal waves can traverse through the liquid zone (Fig. 4.27b). One measures a shadow range for

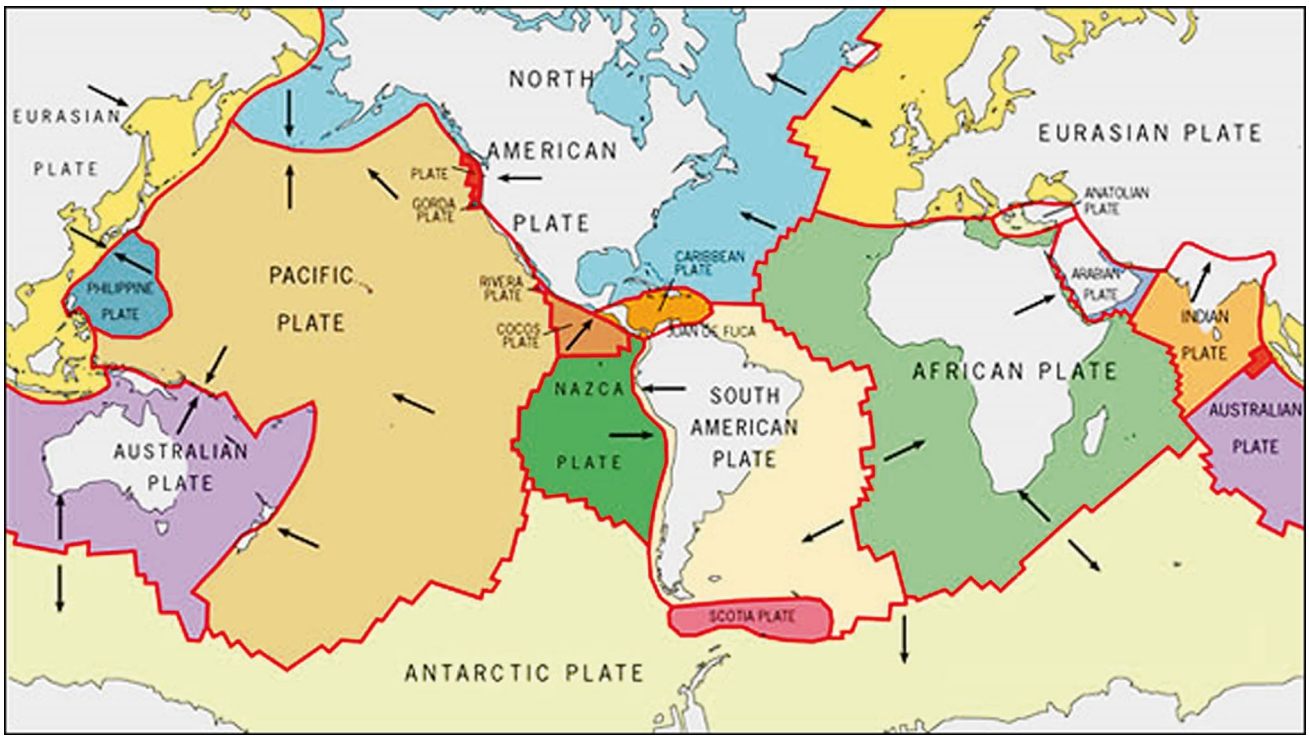


Fig. 4.26 The continental plates (Wikipedia)

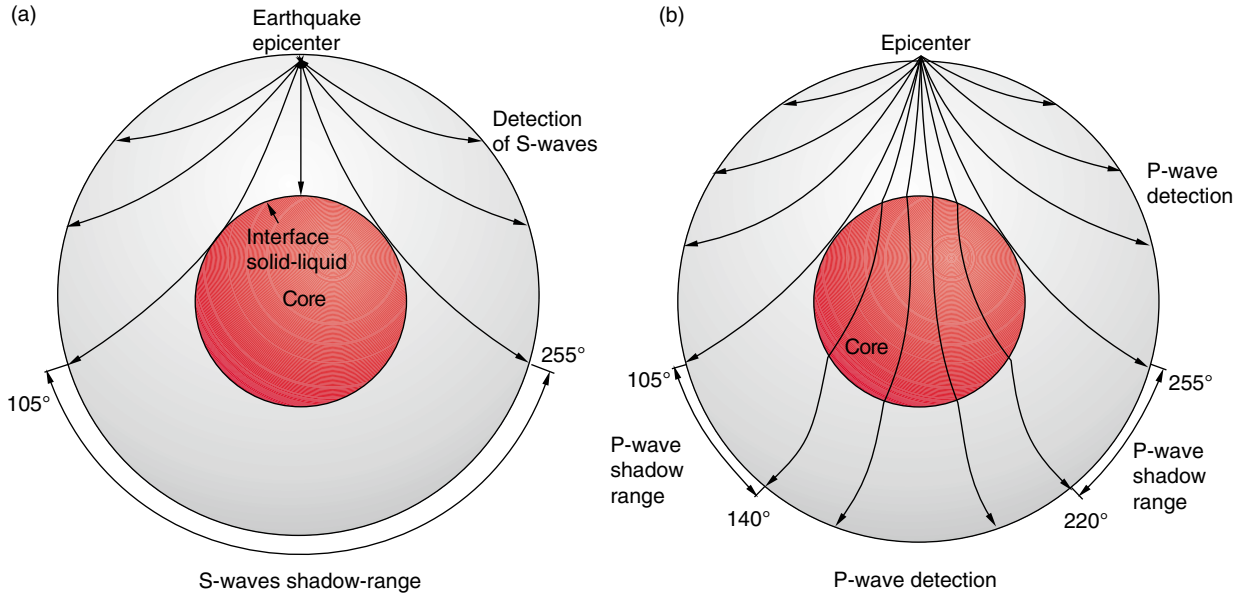


Fig. 4.27 Measurements of the shell structure of the Earth with seismic waves, which are generated by earthquakes. **a** Transverse waves (S-waves or shear waves). **b** Longitudinal waves (P-waves are primary waves)

transverse waves. The angle of this range on the side of the Earth opposite to the source gives information about the size of the liquid zone. The speed of the waves depends on the temperature and the pressure of the material through which the wave passes. It ranges from 2 up to 8 km/s. Seismographic stations record amplitudes and frequencies of these waves and their arrival time, compared to stations at other locations. Longitudinal waves have a higher velocity than transverse waves. They are therefore named *primary* or *p-waves*, while the transverse waves are the *secondary waves* (s-waves).

4.3.3.2 The Earth Atmosphere

The Earth atmosphere has no real outside boundary, because the density decreases nearly exponentially with altitude (Fig. 4.28). The pressure for h = 0 at the Earth surface is $p_0 = 1$ bar = 760 torr. At h = 5 km above ground, the air pressure has decreased to $p = p_0/2$ and for h = 50 km is p = 0, 1 mbar.

The total mass of the atmosphere amounts to 10^{-6} of the Earth mass.

The temperature dependence T(h) is more complex. It is determined by the different heating mechanisms (absorption of Sun radiation and infrared radiation from the ground) and by air convection induced by pressure differences (see Vol. 1. Sect. 7.6). The atmosphere can be divided into five regions (Fig. 4.29):

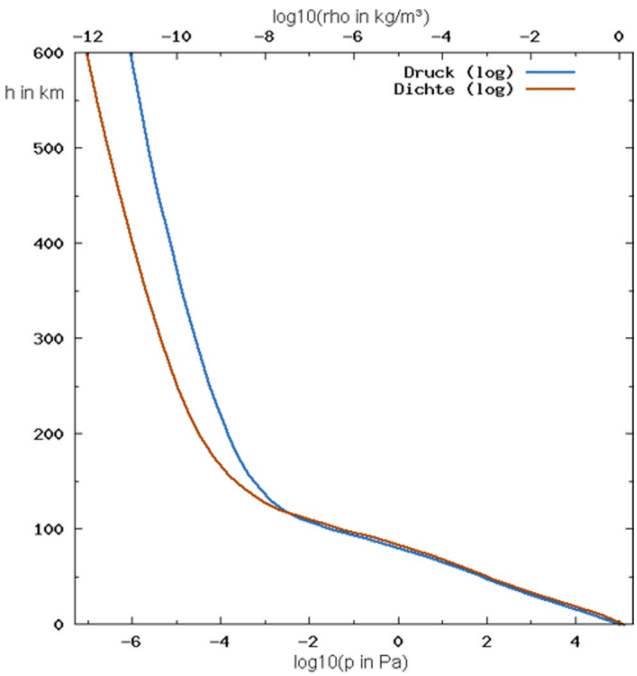


Fig. 4.28 Pressure (blue) and density (brown) of the Earth atmosphere as a function of the altitude h (Wikipedia)

1. *The troposphere* (h = 0–12 km). This is the region where our weather happens. It contains about 75% of the total air mass and 99% of the water vapor. The temperature decreases with increasing altitude because the infrared radiation from the surface of the Earth decreases with h.

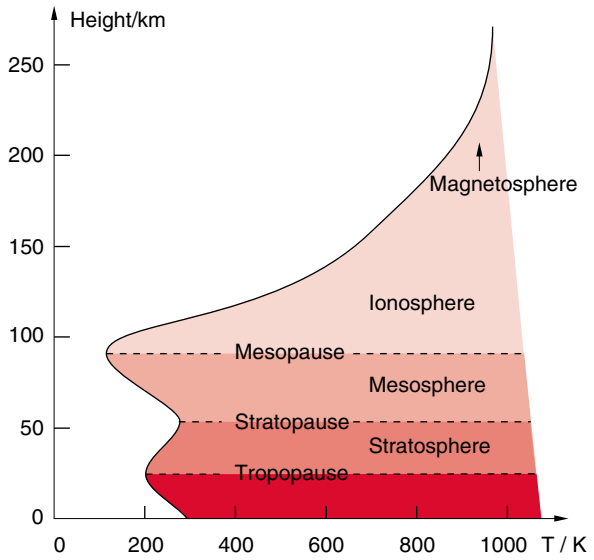


Fig. 4.29 Temperature progression and different layers in the Earth atmosphere

2. *The stratosphere* (12–50 km). Here, the temperature increases with increasing altitude. This increase in the Stratosphere and in the lower part of the Mesosphere around h = 40–70 km is caused by the absorption of ultraviolet sunlight by the ozone layer.
3. The *mesosphere* (around 50–85 km). In the upper part of the mesosphere, the concentration of ozone molecules is smaller, which results in a smaller heating, but the CO_2 molecules emit infrared radiation into space, which represents an energy loss and therefore the temperature decreases again.
4. *The thermosphere* (85–500 km). Here, the temperature rises again. This is due to energetic particles from the Sun which enter the thermosphere and collide with the particles in this range. The temperature is therefore very high (up to 1000–1500 K) and increases with altitude.

These energetic particles have sufficient energy to ionize the atoms and molecules by collisions. Also, the vacuum ultraviolet radiation from the Sun, which can reach this high region with low density contributes to the ionization. The region where the density of ions is maximum is also called *ionosphere*. The motion of the ions is determined by the magnetic field of the Earth.

The ions spiral around the magnetic field lines. Since these lines enter the Earth close to the poles (Fig. 4.30), the ions enter the lower atmosphere in northern regions, collide with the molecules of the atmosphere and the light emission of the excited molecules cause brilliant and impressive light phenomena called *aurora borealis* (Fig. 4.31).

The magnetosphere is that region where the forces caused by the superposition of the magnetic fields of Earth

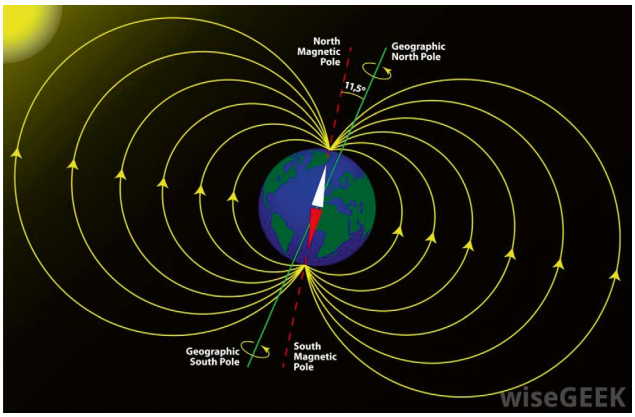


Fig. 4.30 The magnetic field of the Earth (wisegeek.com)

Fig. 4.31 Aurora Borealis (https://www.bing.com/images/search?q=aurora+borealis)

(Fig. 4.30) and Sun and the solar wind are stronger than the gravitational force and therefore govern the motion of charged particles. It extends from the ionosphere (about 500 km altitude) until far outside into space [11]. On the side of the Sun, the solar wind compresses the Earth magnetic field and develops a shock wave Fig. 4.32), while it is expanded on the nightside where it extends up to 1 million km (Fig. 4.33).

In Fig. 4.34, the atmospheres of Venus, Earth and Mars are compared. Furthermore, the variations of day- and night temperatures are given, which depend on the rotation period of the planets and the density of their atmospheres.

The temperatures in the upper layers of the atmospheres of Venus and Mars are smaller than for the Earth because they contain more CO_2, which radiates its excitation energy into space and therefore results in a cooling of the upper atmosphere.

In the lower layers of the atmosphere, the temperature differences between day and night are smaller for Venus and Earth than for Mars because the denser atmosphere absorbs the infrared radiation emitted from the surface and

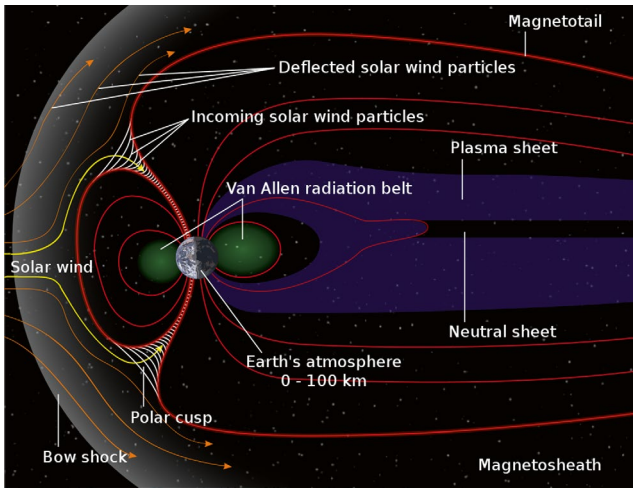


Fig. 4.32 The magnetosphere (NASA Vector: Aaron Kaase, Medium69 Public Domain Wikimedia)

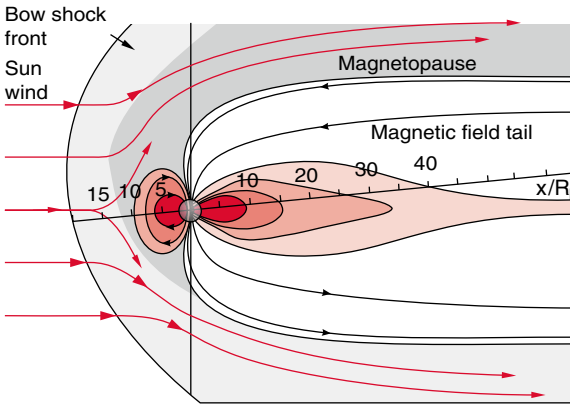


Fig. 4.33 Magnetosphere of the Earth. The numbers give the distances for the direction perpendicular to the magnetic dipole axis in units of the Earth radius R_E [11]

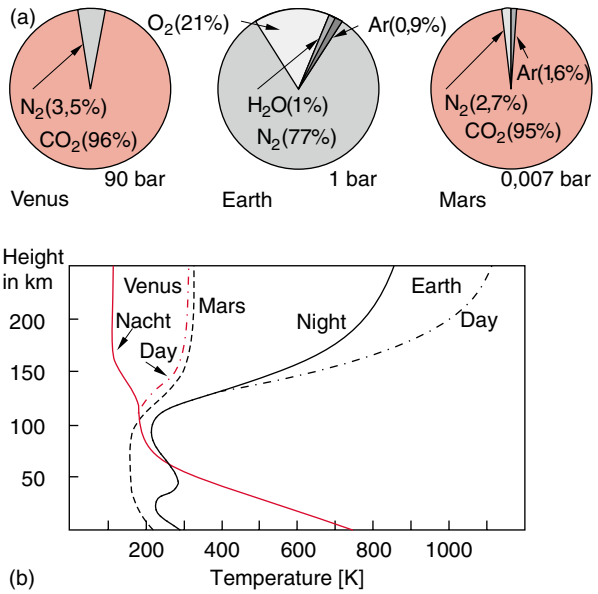


Fig. 4.34 **a** Comparison of the chemical compositions in the atmospheres of Venus, Earth and Mars. **b** Day- and night temperatures (after J. K Beatty: Die Sonne und ihre Planeten. Physik Verlag Weinheim 1995)

therefore partly compensates the missing Sun radiation during night time.

It is interesting to study in more detail the energy balance of the Earth illustrated in Fig. 4.18 in order to esteem the importance of our atmosphere for the energy budget of the Earth [3].

The present Earth atmosphere is essentially determined by biological processes. While the early original atmosphere has been blown away by the solar wind, which was much stronger in these early times, a new atmosphere has been formed by outgasing of the Earth crust and by volcanic actions. This new atmosphere contained only a small concentration of oxygen. After micro-organisms in water had produced sufficient oxygen by photosynthesis of CO_2 and water, living organisms could spread out from water to solid surfaces. The present oxygen concentration is mainly determined by photosynthesis (assimilation) of green plants, mainly forests. Since these green plants are missing on Venus and Mars, their atmospheres contain only small traces of oxygen.

4.3.3.3 The Earth Moon

The origin of the Earth moon was for a long time controversially discussed. Recent detailed investigations of rocks from the moon (brought back by astronauts) and refined simulation models have confirmed the hypothesis that the moon was created by the impact of a large celestial body (probably an asteroid) on the Earth, soon after the Earth had been formed. This impact has hurled out a sizeable fraction of the Earth which became the moon. This explanation is an agreement with all experimental facts obtained so far, as will be shown in the following [5].

While for all moons of other planets, the orbits of moon and planet lie in the same plane, which indicates that planet and moon have been formed simultaneously, the orbit of the Earth moon is inclined against the ecliptic by an angle of 5°9′ (Fig. 4.35). Since the equator plane of the Earth is tilted against the ecliptic by the angle 23.5°, the angle between moon orbit and equator plane of the Earth amounts to 28.5°. The moon appears for an observer on Earth within the angular range of ±28.5° against the equator plane of the Earth.

The mass ratio $m_{\text{moon}}/m_{\text{earth}} = 0.012$ is larger than $m_{\text{moon}}/m_{\text{planet}}$ for all other moons in the solar system.

The moon is, after the Sun, the second brightest object on the sky.

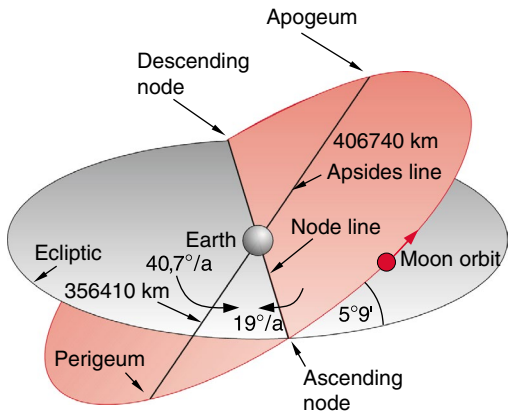


Fig. 4.35 Inclination of the orbital plane of the moon (red plane) against the ecliptic and moon orbit with the maximum (Apogeum) and the minimum (Perigeum) distance from the Earth (after A Unsöld, B. Baschek: Der Neue Kosmos (Springer Heidelberg 1991))

The relative positions of Sun, Moon and Earth determine the moon phases, which define the periodically varying illuminated fraction of the moon surface (Fig. 4.35). At the full moon phase, the moon is in opposition to the Sun, at new moon it is in conjunction, i.e. its unilluminated part points towards the Earth.

A lunar eclipse can only occur at the full moon phase, i.e. when the moon is located on the opposite side of the Sun and the Earth is on the connecting line Sun-moon. A solar eclipse occurs only at the new moon phase, i.e. when the moon stands between Sun and Earth (Fig. 4.36).

The question is now, why is there not every month a lunar or solar eclipse?

If the moon orbital plane would coincide with the ecliptic, there would be indeed every 29.5 days a solar eclipse, and, delayed by 14.7 days, a lunar eclipse. However, since the moon orbital plane is inclined against the eclipse, such constellations occur much more seldom, namely exactly at such times when the moon passes the intersection line between the two planes (nodal line) which must simultaneously coincide with the line Sun-Earth.

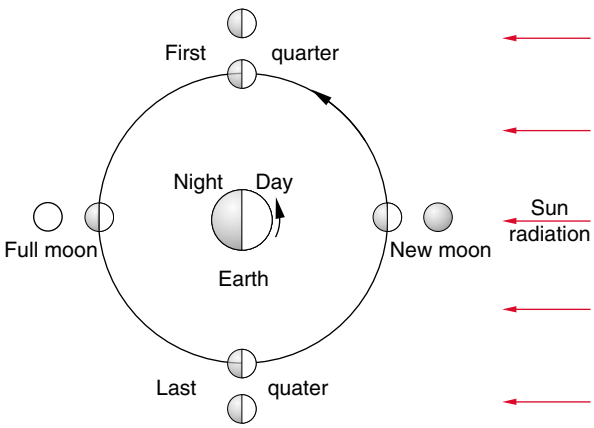


Fig. 4.36 Explanation of the moon phases (After A. Unsöld, B. Baschek. Der Neue Kosmos (Springer Heidelberg 1995)

The orbital angular momentum of the moon relative to the Earth is not constant in time because of the perturbation by the Sun. The nodal line moves by 19.4°/year (full period is 18.6 years) into opposite direction as the moon revolves around the Earth (Fig. 4.33). Also, the line of apsides (the line between the *Apogeum* (the point nearest to the Earth) and the *Perigeum* (the point farthest from the Earth) turns by 40.7°/year (period is 8.85 years) in the same direction as the moon revolution.

The exact prediction of lunar or solar eclipses therefore demands a detailed extensive calculation, which takes into account all these motions. However, there exist an empirical rule, known already in ancient times, that the mean time interval between successive eclipses is 18 years. This makes approximate predictions much more simple.

While a total lunar eclipse can be seen from all points on the Earth for which the moon appears above the horizon (Fig. 4.37b) a total solar eclipse is visible only in a small stripe on the Earth (Fig. 4.37a). The reason is because the umbra shadow of the Earth is larger than the diameter of the moon, while the umbra shadow of the moon is with 270 km diameter much smaller than the Earth diameter. In Fig. 4.38, photos of a total solar and a lunar eclipse are shown.

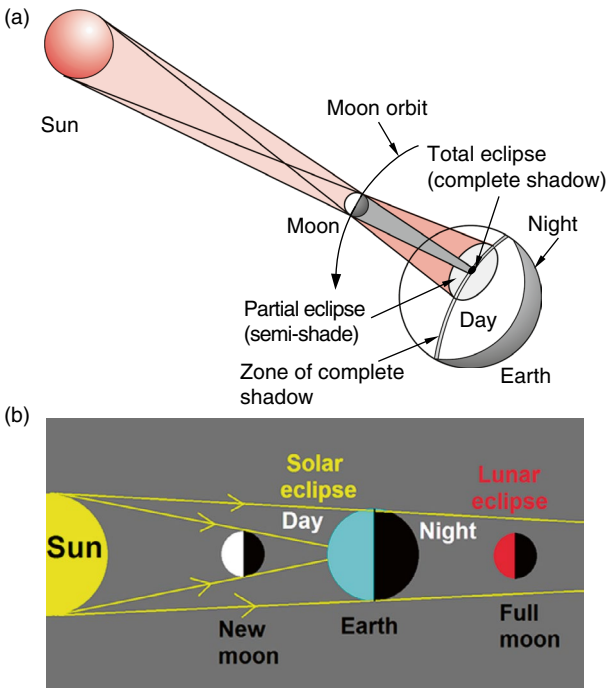


Fig. 4.37 **a** Explanation of Sun eclipse (A. Unsöld et al.) **b** vIllustration of solar and lunar eclipses (Wikipedia)

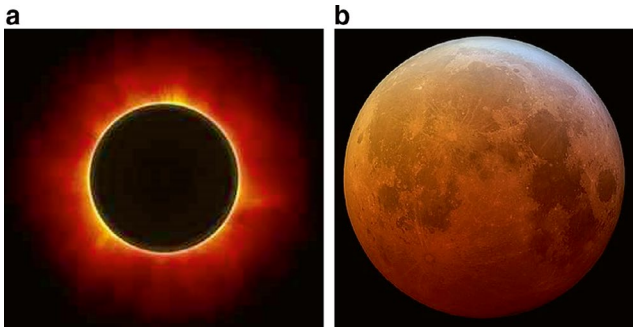


Fig. 4.38 **a** Total solar eclipse, **b** total lunar eclipse January 2019 (Wikipedia)

Similar to the situation for the planets, we distinguish also for the moon between synodal and sidereal revolution periods around the Earth. The sidereal month is the period of 27.3 days, the moon takes for a revolution of 360° relative to the stars. The synodal month is the time between to equal phases of the moon. During one revolution of the moon the Earth has moved on its way around the Sun by 30°. The moon has, therefore, to proceed by $360° + 30° = 390°$. Similar to Eq. (4.21a), we obtain

$$\frac{1}{T^{\text{syn}}_{\text{Mond}}} = \frac{1}{T^{\text{sid}}_{\text{Mond}}} - \frac{1}{T^{\text{sid}}_{\text{E}}}, \tag{4.34}$$

where $T^{\text{sid}}_{\text{E}}$ is the sidereal period of the Earth. From $T^{\text{sid}}_{\text{Mond}} = 27.3\,\text{d}$ and $T^{\text{sid}}_{\text{E}} = 365.2\,\text{d}$, we obtain $T^{\text{syn}}_{\text{Mond}} = 29.5\,\text{d}$.

Earth and moon move around their common center of gravity S, which lies still inside the Earth, 4670 km from the center.

Due to the gravitational perturbation by the Sun, the orbit of the moon is no longer a pure Kepler ellipse with constant orbit elements. The mean major axis of the moon orbit is $a = 384.400\,\text{km}$, which is the mean distance between Earth and moon.

The eccentricity of the moon orbit is $\varepsilon = 0.055$. The minimum and maximum distances from the Earth are not exactly equal to $a(1 \pm \varepsilon)$ because of the perturbations, but

$$\begin{aligned} r_{\text{min}} &= 356.410\,\text{km}, \\ r_{\text{max}} &= 406.740\,\text{km}. \end{aligned}$$

The distance from a selected point on Earth to the retroreflector on the moon can nowadays be measured with an accuracy of better than 1 cm!! by measuring the time, a short laser pulse needs to travel from a station on Earth to the reflector which had been installed on the moon by the astronauts (see Sect. 4.4).

The moon always shows the same side to the Earth, i.e. its rotational period is equal to its revolution period (bound rotation). This has not been always the case. There are proofs that the moon has rotated faster in former times. Due to the attraction of the moon by the Earth the moon crust deformes, completely similar as the tides on Earth. As long as the moon rotated, the tide wave moved around the moon, which decelerates the rotation due to friction, until the rotation period equals its synodal revolution period. Then the friction stops since now the moon always shows the same side to the Earth and the deformation on the moon remains stationary.

Exactly the same happens with the Earth. Its rotation is slowed down due to the tide friction. The interesting point is that the angular momentum (spin) of the rotation decreases. Since the total angular momentum must be conserved, (there are no external torques acting on the Earth-Moon system), the orbital angular momentum $L = r \times \mu \cdot v$ of the system must increase, where $\mu = m_E m_M/(m_E + m_M)$ is the reduced mass.

Since the orbital velocity v of the moon cannot increase (this would mean an increase of its kinetic energy), its distance r from the Earth must increase. The moon must have been in former times closer to the Earth. This fact also supports the theory that the moon is generated from a fraction of the earth after a heavy collision of a large planetoid with the Earth.

▶ If a moon is closer to the planet than the synchronous distance r_s (this is that distance where the revolution period of the moon is equal to the rotational period of the planet, the velocity of the moon will decrease due to tide friction and the moon will finally fall onto the planet. If $r > r_s$, the moon will be accelerated while the rotational energy of the planet will decrease and the distance between moon and planet will increase.

For $r < r_s$, the moon revolves faster than the rotation of the planet. The gravitational force between the distorted sphere of the moon will slow down its velocity. For $r > r_s$, the rotation of the planet is faster than the revolution of the moon and the force accelerates the moon.

For the Earth moon is $r_s = 42.000\,\text{km}$ (geostationary orbit). When the moon was hurled out from the Earth, its initial kinetic energy must have been large enough to bring it beyond the distance r_s.

Much information about the chemical composition of the Earth moon and details of its surface structure have been available after the unmanned and manned moon missions,

Fig. 4.39 The first man (Neil Armstrong) on the moon with the lunar vehicle Eagle (NASA)

where many mineral samples have been brought back to Earth and were carefully analyzed (Fig. 4.39).

4.3.4 Mars

Since Mars is of all upper planets, the one which is closest to the Earth it has activated the particular interest of mankind. One reason is its reddish-yellow color, which the human phantasy related to fire and war. It was therefore named already in ancient times as *Mars, the God of war*.

Its equator diameter is with 6794.4 km about half of the Earth diameter. Its mass is $m_M = 0.11 m_E$. Its sidereal rotational period is $T^{sid} = 24^h37^m22.7^s$, its synodal period is $T^{syn} = 24^h33^m$.

One Mars day has therefore about the same duration as the Earth day. The inclination of the rotation axis is about the same as for the Earth. Therefore, on Mars appear similar to the situation on Earth annual periodic temperature variations, in particular at the poles. Indeed, observations with telescopes on Earth show white pole caps, which vary in size with an annual period. During the wintertime, they are large and decrease in size during summer time.

In 1817, G. V. Schiparelli drew a detailed map of the Mars surface, which was based on the compilation of observations of many astronomers. On this map, he drew many dark lines, which he named "*Canali*". This was exploited by popular science papers as manmade channels built by intelligent Mars residents [6]. Mars observations with large telescopes could not confirm these channels which induced a hot controversy among different observers. Later on, it turned out that huge canyons (Fig. 4.40) had been misinterpreted as man-made channels.

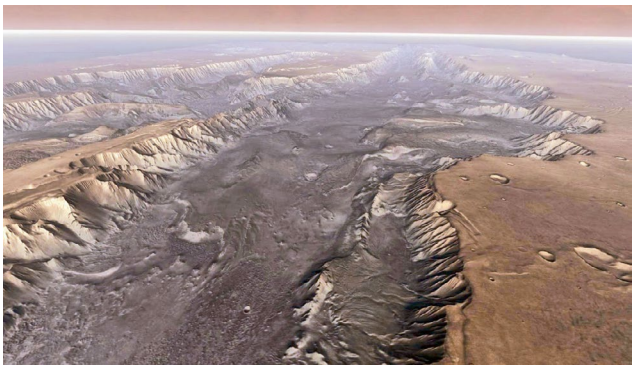


Fig. 4.40 Canyon on Mars (NASA)

The atmosphere of Mars consists of 95.3% CO_2, 2.6% N_2, 1.9% argon and small percentages of O_2 and water vapor, which comes mainly from the northern pol cap which is covered by a layer of water ice and CO_2 ice, while the southern pole cap is covered only by CO_2 ice, which melts in the southern summer, because the melting temperature of CO_2 is only $T = 216\,K$ at $p = 9\,mbar$. The total pressure of the Mars atmosphere is $610\,Pa = 6 \cdot 10^{-3}\,bar = 6\,mbar$, which is less than 1% of the Earth atmospheric pressure [6].

Because of the thin atmosphere, the diurnal and annual variations of the temperature are essentially determined by the solar radiation, since the storage of heat and the redistribution by convection in the atmosphere plays only a minor role. This is different from the situation on Earth. The temperature on the Mars equator sinks in the Mars summer bevor sunrise to $-100\,°C$ while it rises up to $20\,°C$ at noon.

More detailed observation of the Mars surface could be performed by unmanned space probes. Already in 1971, the American space probe Mariner 9 was navigated into an orbit around Mars and sent excellent sharp photos to Earth. Particular important photos were made by the Viking mission (1975–1976), where a probe landed on the mars and took very detailed photos of the surroundings of the landing place (Fig. 4.41). Using the experiences with the former missions, the space probe Phoenix was started in 2007, which was equipped with many measuring and analyzing devices. It landed on Mars in 2008, drilled 0.5 m deep holes into the Mars surface and analyzed the obtained samples [6]. The analysis of the drill cores should give information whether there had been water on Mars. The results confirmed that there is still water on Mars but no signs of living cells had been found. A recent analysis of a meteorite found on Earth, which had been ejected by mars from its deeper interior after the impact of a larger body, has found traces of early primitive forms of living cells. These results are, however, controversially discussed. ($p_0 = 9\,mbar$).

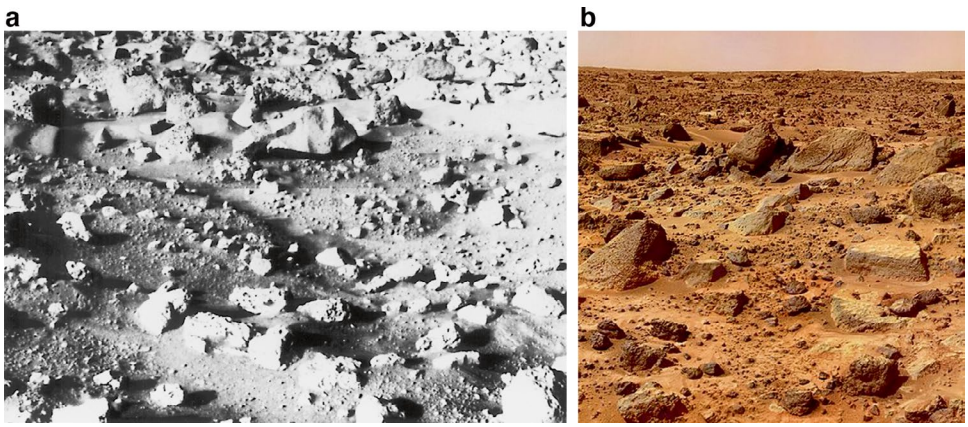


Fig. 4.41 **a** Mars surface around the landing place of Viking 2 (NASA). **b** Section of Mars surface, covered with rocks, photographed 1997 by Mars Pathfinder (NASA)

The most extensive exploration of the Mars surface was performed by the Mars probe *Curiosity,* which landed in the Gale-crater of Mars [4, 5]. It carried a wheeled rover with 17 cameras and 10 analyzing devices (Fig. 4.42). It can move over the Mars surface and has made photos of its surroundings. With a laser drilling system, it has digged out many samples for the investigation of the crust of the mars surface. The analysis of these samples with a mass spectrometer on board allowed the determination of the atomic and molecular constituents. With X-ray diffraction devices, it was possible to determine the crystallographic structure of minerals in the samples.

The solid surface of Mars is polymorphic with regard to its structure and can be ranged between the surfaces of the Earth and the Earth moon [7].

The Mars probes could also measure the atmospheric temperature as a function of height h (Fig. 4.43). When the results were compared with model calculations it turned out that dust played an important role for the temperature distribution.

The highest mountain on Mars is the volcano Mars Olympus. It is with a height of 26 km above the plains in the surroundings (22 km above the mean Mars surface) and with a base diameter of 600 km the largest mountain in the whole solar system (Fig. 4.44). The Mars crust must be thick enough to carry such a large mountain.

About 40% of the Mars surface are covered with impact craters. There are large planes covered with volcanic lava, which have overcast earlier impact craters [6, 7].

Some in particular impressive structures on the Mars surface are huge canyons (Fig. 4.40), where the largest is 2700 km long, more than 100 km broad and up to 6 km deep. This should be compared with the great canyon in Arizona, which is 35 km long and 1.6 km deep and

Fig. 4.42 Self-portrait of Mars Probe Curiosity (NASA)

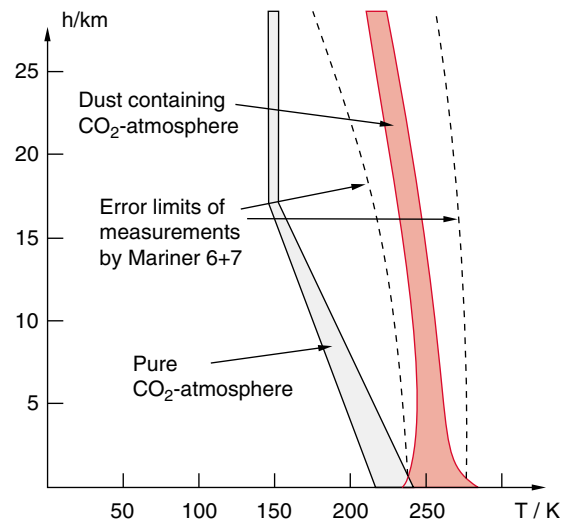


Fig. 4.43 Simulated temperature progression as the function of height in the atmosphere of Mars, with pure CO_2-atmosphere (grey) and CO_2-atmosphere mixed with dust (red). The dashed lines give the error limits of measurements by the Mariner probes (after F. H. Shu: The Physical Universe (University Science Books, Mill Valley 1982)

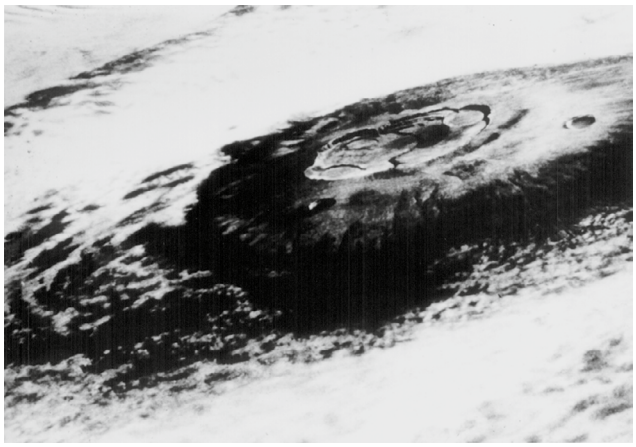


Fig. 4.44 The volcano Mons Olympus is with 25 km height the largest volcano mountain in the solar system (NASA)

therefore a dwarf compared to the canyon on Mars. There are no unambiguous insights about the origin of these canyons. They have been probably formed by liquid floods (water or lava) from volcanic mountains into the planes.

In March 2021 a new, more sophisticated Mars rover, called *Perseverance*, landed on the Mars surface (Fig. 4.44). Its microphone system recorded a wind noise, which was surprising because the air density on Mars is only 1% of that of the Earth atmosphere. This rover could drill up to 1 m into the Mars surface and could measure with a spectrometer the composition of the different layers in the upper crust of the surface (Fig. 4.45).

Mars has two small moons:

- Phobos is an irregularly formed body that looks like a potato (Fig. 4.46) with a major diameter of 27 km ($26.6 \times 22.2 \times 18.6$ km) and a mass of 10^{16} kg $= 1.5 \cdot 10^{-9}\, m_E$, which orbits Mars at a distance of 9.380 km with a period of 7 h 39 min.
- Deimos ($15 \times 12.4 \times 10.8$ km), m $= 2 \cdot 10^{15}$ kg $= 4 \cdot 10^{-10}\, m_E$. Its surface is very dark (albedo A = 0.05). It orbits around Mars at a distance of 23.500 km with a period of 1.262 d.

Phobos and Deimos consist probably, like many other planetoids, of carbon-rich chondrites with a density of 2 kg/dm^3.

4.4 The Outer Planets

At the time of Galilei, only Jupiter and Saturn were known, of all outer planets. Galilei, who was the first to observe the sky with his newly developed telescope, discovered the 4 "Galilean moons" of Jupiter, (named Io, Europe, Ganymed and Calisto) and measured their periods. This discovery represented the first proof that the planets were not fixed on shells of glass as had been assumed before because Galilei concluded that the moons would break the glass shells when circulating around Jupiter.

The planet Uranus was discovered in 1781 by the amateur astronomer William Herschel. From observed perturbations of the Uranus orbit, John Couch Adams from Cambridge and Urbain Jean Joseph le Verrier from Paris calculated independently the position of a new planet, which was 1846 indeed discovered by Johann Gottfried Galle and was named Neptune.

After the downgrading of Pluto to a dwarf planet, there are now eight planets in our solar system. Neptune is the outermost planet in the solar system, However, beyond the Neptune orbit, there are still a large number of small celestial bodies, which are mainly located in a range between 30 and 59 AU from the Sun, which is called the Kuiper-belt (Gerard Peter Kuiper 1906–1973, Dutch-American Astronomer) [10].

The observations of a series of comets can be best explained under the assumption that they originate from a

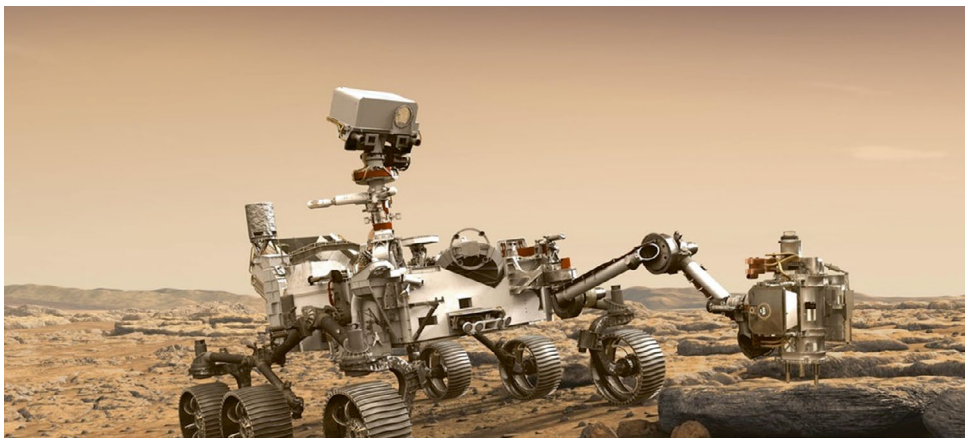


Fig. 4.45 Self-portrait of the new Mars rover *Perseverance* (NASA)

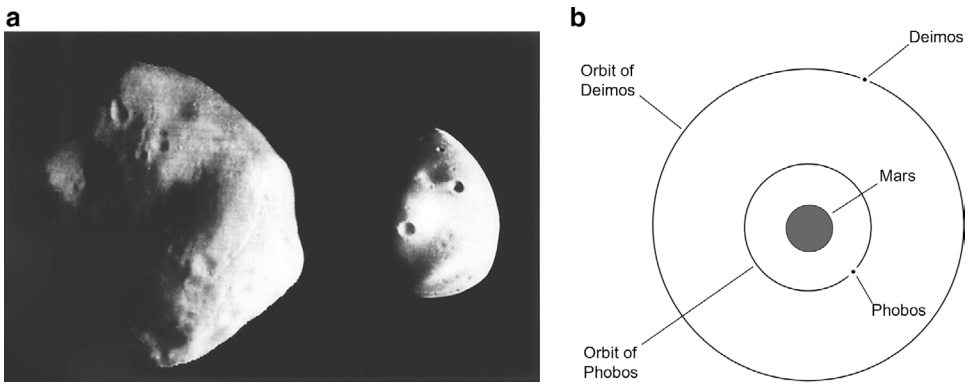


Fig. 4.46 **a** The two moons of Mars: Phobos (left) and Deimos (right). The photo was taken 1977 by Viking Orbiter (NASA). **b** Their orbits around Mars (Von Muskid-Eigenes Werk, CC BY-SA 3.0. https://commons.wikimedia.org/w/index.php?curid=33575180)

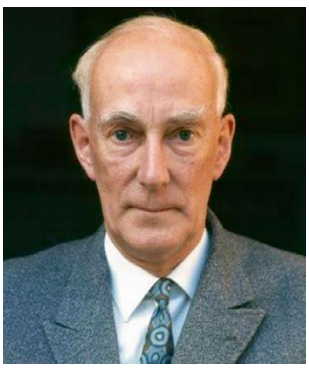


Fig. 4.47 Jan Hendrik Oort (Wikipedia)

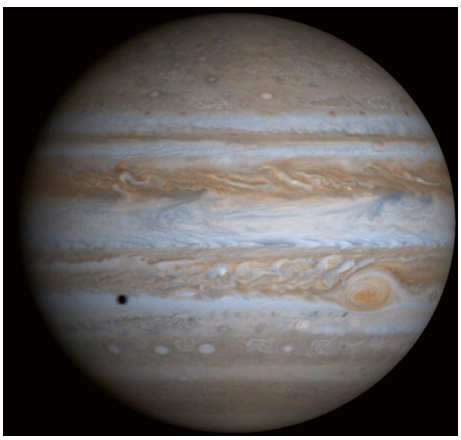


Fig. 4.48 Cloud structure of Jupiter (Von NASA/JPL/University of Arizona)

far distant extended range (about 50.000 AU ≈ 1 Ly) called the Oort's cloud which contains many thousand cometes. It is named after the Astronomer Jan Hendrik Oort (1900–1992) (Fig. 4.47).

Neighboring stars passing close by may have changed the orbits of bodies in the Oort's cloud by gravitational interaction and deflected them into the solar system, where they perform elongated very eccentric elliptical orbits, thus coming close to the Sun. The second possibility of their fate is that they are thrown out of the solar system by these passing stars.

It should be, however, emphasized that the existence of the Oort's cloud is up to now pure speculation and has to be confirmed by further experimental proof (Fig. 4.47).

4.4.1 Jupiter and Its Moons

Jupiter is by far the largest planet in our solar system. Its equator diameter is 142.796 km, its mass is $m_{\text{Jupiter}} = 317.9\, m_{\text{E}} = (1/1047) M_{\odot}$. It has a dense cloud structure with stripes parallel to the equator (Fig. 4.48).

In particular impressive is the big red spot, which had been already discovered 1665 by G. D. Cassini. It represents a huge long living atmospheric vortex, which exists over several centuries. Much more detailed photos have been shot by the various space probes which passed close by Fig. 4.49. These space probes have brought a wealth of detailed new information about the structure of Jupiter's atmosphere, which shows cloud stripes, vortices and convection zones, exhibiting large velocity gradients.

Jupiter rotates extremely fast, not as a rigid body but shows a differential rotation, i.e. the angular velocity $\omega(\varphi)$ depends on the latitude φ. The largest value of ω is observed for $\varphi = 0$ at the equator, where the rotational period is T = 9 h 50 min, while the regions around the poles rotate more slowly. This differential rotation causes the observed large velocity gradients between neighboring circles of latitude and leads to the formation of huge vortices.

Spectroscopic measurements and in particular the results of the space probes Pioneer and Voyager have explored the atomic and molecular composition of the atmosphere (Table 4.5). Besides the major part of 90% H_2 molecules and 10% Helium atoms, traces of Methane CH_4, Ammonia NH_3, acetylene C_2H_2, ethane C_2H_6, water H_2O, Hydrogen cyanide HCN and Phosphine PH_3 were found. The highest cloud layers have a mean temperature of about 150 K and contain frozen NH_3 crystals.

Fig. 4.49 The large red spot with Karman Vortex street on the Jupiter surface. The photo was taken by the space probe Voyager 2 from a distance of 6 Million km (NASA)

Table 4.5 Relative abundance of different molecules in the atmospheres of Jupiter, Saturn and the Sun for a model atmosphere with solar composition

Atom/molecule	Jupiter	Saturn	Sun
$H + H_2$	0.90	> 0.94	0.89
He	0.10	< 0.06	0.11
H_2O	$1 \cdot 10^{-6}$	–	$1 \cdot 10^{-3}$
CH_4	$7 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$
NH_3	$2 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$

The analysis of all observations brought the result that Jupiter consists mainly of solar material, i.e. the same composition of 90% H atoms and 10% He atoms as in the Sun.

Jupiter shows a much more pronounced oblateness $O = (R_{equ} - R_{pol})/R_{equ} = 0.067$ than the Earth. This is partly due to its fast rotation and mainly to the fact that it consists of gases instead of solid material. If Jupiter would consist solely of gas the oblateness would be even larger. Therefore, one can conclude that the core of Jupiter should consist of heavy solid material. Probable this is, like in the Earth, an iron-nickel core which is encased by a layer of metallic solid hydrogen (Fig. 4.50). Because of the extreme pressure (about $3 \cdot 10^6$ bar) in the interior of Jupiter, molecular hydrogen H_2 will be dissociated at a temperature of $T > 10^4$ K and most of the H-atoms are ionized.

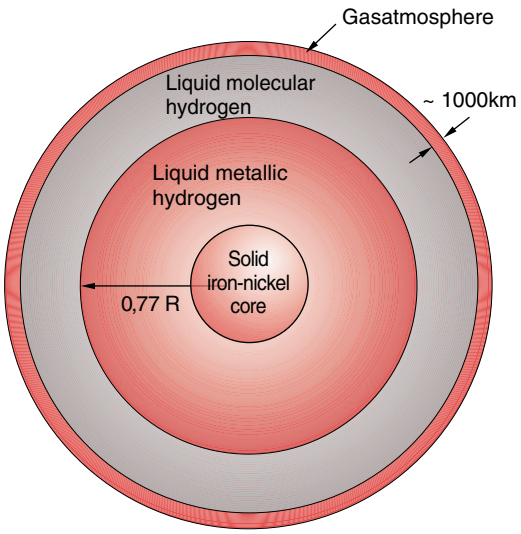


Fig. 4.50 Layer Structure of Jupiter

From the equation of state $p = N \cdot k \cdot T$, the mean density of the H-atoms is $N = 2 \cdot 10^{30}/\mathrm{m}^3$, which gives a mean mutual distance between the atoms of $\overline{d} = 1.3 \cdot 10^{-10}$ m. The Fermi-energy of the electrons $E_F \propto \hbar^2/(m_e\, \overline{d}^2)$ is therefore larger than their binding energy, and the free electrons move quasi-free in a lattice of protons, like in a crystalline solid. This state is called *metallic*.

At a radius of 0.77 R_{jup} at a pressure of $3 \cdot 10^{11}$ Pa and a density of $\varrho = 10^3\ \mathrm{kg/m}^3$, the phase transition from liquid metallic hydrogen to liquid molecular hydrogen takes place (Fig. 4.50). The subsequent layer up to the atmosphere consists mainly of liquid molecular hydrogen. The thickness of the gaseous atmosphere amounts to about 1000 km. The cloud layer is, however, only 140 km thick, which is 0.2% of the Jupiter radius R_{Jup}.

A further hint to this metallic core of iron, nickel and metallic hydrogen is the magnetic field of Jupiter, which is about 19.000 times larger than the Earth magnetic field. Its direction is inclined against the rotation axis by 9.5° (Fig. 4.51).

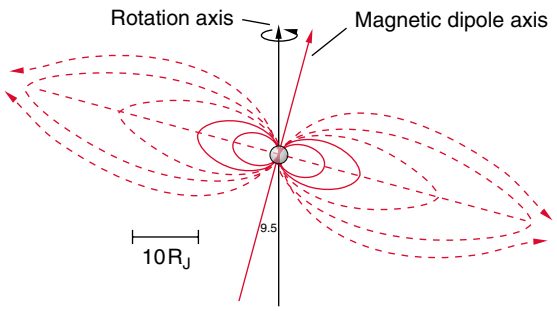


Fig. 4.51 The magnetosphere of Jupiter constructed after Pioneer 10 measurements in 1973. It reaches out into space up to 100 Jupiter radii (NASA)

Fig. 4.52 The four Galilean Moons with increasing distance from Jupiter: Io, Europe, Ganymed and Kallisto (from left to right) (NASA)

Because of the strong magnetic field, corotating with Jupiter, charged particles, in particular electrons are accelerated and emit synchrotron radiation. The maximum of this radiation lies in the radio-frequency range and can be detected on Earth as radio radiation from Jupiter.

Jupiter emits about twice as much energy as it receives from the Sun. Although the energy balance it restored by convection currents from the hotter inner part to the surface, the total energy of Jupiter decreases in time. It therefore either cools down or it contracts slowly to compensate the net energy loss by gravitational energy.

Up to now 79 moons of Jupiter have been observed. The first 4 moons discovered by Galilei when he directed his new telescope towards Jupiter are with increasing distance from Jupiter (Fig. 4.52): *Io, Europe, Ganymed* and *Callisto*.

Figure 4.53 shows an interesting photo, where the moons Io and Europe can be seen in front of Jupiter.

The size of the Galilean moons is comparable to that of the Earth moon. They orbit around Jupiter within its magneto-sphere and are therefore bombarded with high energy ions, which had been captured by the magnetic field of Jupiter. This can contribute to the erosion of the moon surfaces.

The moon Io shows active volcanism (Fig. 4.54). During the passage of Voyager, several volcanic eruptions could be observed. The surface of Io contains sulfur and has therefore a yellow-red color.

A large number of Caldera (volcanic cones with collapsed central part) cover the surface of Io, which appears therefore pockmarked. Due to the volcanic eruptions, the surface of Io is covered in the course of time more and more with a layer of volcanic ash, which covers up the impact craters.

The energy for heating up Io stems probably from periodic compressions caused by the periodic

Fig. 4.53 The moons Io and Europe in front of the Jupiter surface. The picture was taken by Voyager 2 at a distance of 20 Million km (NASA)

Fig. 4.54 Eruption of a Volcano (blue cloud) at the right side of Io (NASA)

Table 4.6 Characteristic data of the four Galilean moons

Moon	Mean distance from Jupiter/km	Sid. period	Radius	Mass	Density
		Per days	km	kg	kg/dm^3
Io	412.600	1.769	1816	$8.9 \cdot 10^{22}$	3.55
Europa	670.900	3.551	1536	$4.9 \cdot 10^{22}$	3.04
Ganymed	1.070.000	7.155	2638	$1.5 \cdot 10^{23}$	1.93
Callisto	1.880.000	16.689	2410	$1.1 \cdot 10^{23}$	1.81

gravitational force acting on Io on its elliptical orbit (tides). The deviation from a circular orbit is caused by resonances between the periods of the different moons ($T_{Io} : T_{Europa} : T_{Ganymed} = 1 : 2 : 4$), which enforce perturbations of the moon orbits. This causes a periodic variation of the gravitational force acting on Io and induces the tide-deformation of the spherical solid of Io (see Problem 4.7).

Contrary to *Io*, the moon *Europa* shows a bright low-contrast surface with a network of lines. Larger craters (with a diameter > 50 km) are missing. This hints to a lower age ($\approx 10^8$ y) of Europa's surface, which is covered with an ice layer. The albedo of Europa is therefore markedly higher than that of Io.

Also, Ganymed is covered with ice, which is, however, partly overcast by dark areas, which have been produced by volcanic eruptions.

The surface of the outer moon Callisto shows much more impact craters than that of the other three moons.

In Table 4.6, the characteristic features of the Galilean moons are compiled.

Jupiter has to our present knowledge altogether 79 moons and a faint ring system with a thickness of 30 km (!), which is located within the range 1.71–1.81 $R_{Jupiter}$ [14]. It consists of many small particles with diameters between some micrometers to some meters which are orbiting around Jupiter on circular orbits (Fig. 4.55).

This ring system was first discovered by Pioneer 11 in the year 1974 and later on more closely investigated by Voyager. It exhibits a radial structure, which is caused by resonances between the orbiting ring particles and moons of Jupiter (Fig. 4.55).

4.4.2 Saturn

With a diameter of 120.000 km and a mass of 95 mE, Saturn is the second largest planet of our planetary system. Its mean density is $\overline{\varrho} = 0.7\,kg/dm^3$, which is smaller than the density of water. Its atmosphere consists to 93% of hydrogen, 7% of Helium and small amounts of methane and ammonia. The mass of its solid iron-nickel core amounts to about 25% of the Saturn mass. Its rotational period is Trot = 10 h 30 min. Its rotation axis is inclined against its orbital plane by 26.7°. This implies that after about 15 years one can see alternately the north pole or the south pole.

The striking feature of Saturn is its ring system (Fig. 4.56), which had been already discovered in 1610 by Galilei. At a new observation 2 years later, Galilei noted to his great surprise that the ring system had disappeared.

Today we understand the reason for this effect: The ring system lies in the equatorial plane of Saturn. Due to the large inclination angle of the rotation axis, there are positions where we look from the Earth exactly into the equatorial plane. Since the ring system has a very small thickness the rings cannot be seen at this position with telescopes of moderate size.

Up to now 62 moons of Saturn have been discovered. The largest is Titan with a mass of $m = 1.35 \cdot 10^{23}\,kg \approx 2\,m_{Moon}$ twice as much as that of the Earth moon. Nine moons had been already observed from the Earth, but the other 53 had been only discovered by several planetary missions with space probes (Voyager and Cassini missions). Some of the moons orbit around Saturn retrograde,

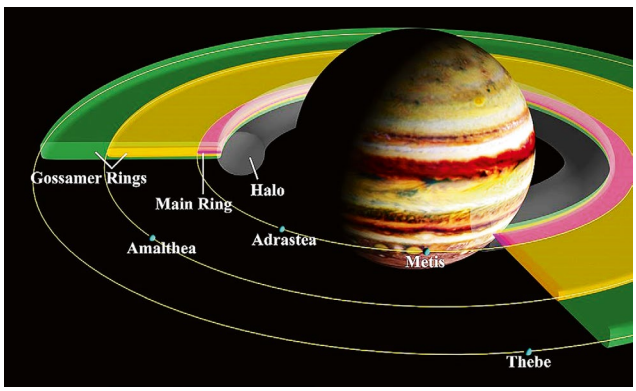


Fig. 4.55 A schema of Jupiter's ring system showing the four main components. For simplicity, the moons Metis and Adrastea are depicted as sharing their orbit. (In reality, Metis is very slightly closer to Jupiter). (By NASA/JPL/Cornell University)

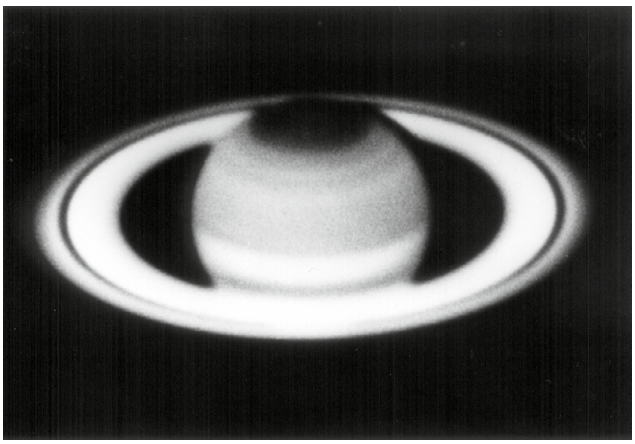


Fig. 4.56 Photo of Saturn with its ring system, taken through a telescope from the Earth

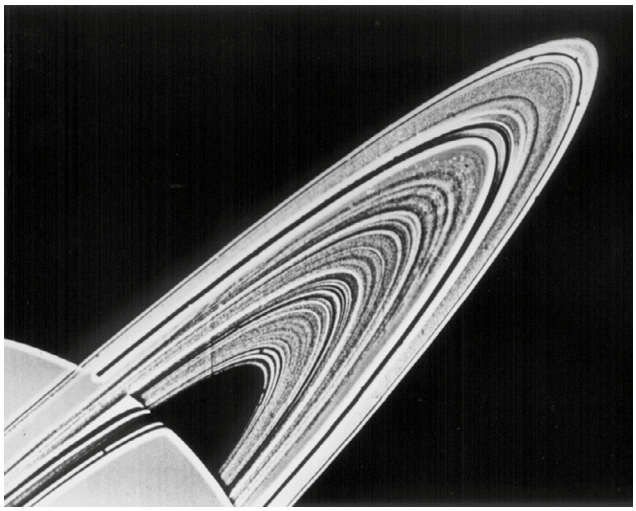


Fig. 4.57 Structures in the ring system of Saturn, photographed by the space probe Voyager (NASA)

i.e. opposite to the rotation direction of Saturn. They cannot have been originated together with Saturn but must have been captured from outside the Saturn moon system.

The inner structure of Saturn is similar to that of Jupiter. Its atmosphere is less structured than the Jupiter atmosphere.

The ring system is distinctively structured (Fig. 4.57) and consists of several subsystems, which are denoted by capital letters. The inner D-ring extends from $r = 67.000$ km (this is only slightly larger than the radius of Saturn) up to $r = 74.000$ km. The outer part of the ring system (E-ring) reaches up to $r = 480.000$ km. Its diameter is therefore about 8 times the diameter of Saturn. The thickness of the ring system is with 0.4 km astonishingly small.

The Voyager space probe discovered radial structures in the B-ring, which are called "*spokes*". The gap between the A and B-ring is called the *Cassini division*. It is not completely empty but contains darker particles with a much smaller reflectivity. It can be already observed with Earth-bound telescopes of medium size (15–20 cm objective diameter).

The Cassini space probe showed much more detailed structures in the ring system (Fig. 4.58).

The particle sizes in the ring system range from tiny dust particles with diameters of a few μm to large boulders with up to 10 m diameters, where the maximum of the size distribution lies between 0.1 and 1 m. Composition and size distribution of the ring particles can be obtained from the spectral reflection and the ratio of forward and backward scattering intensities. The small particles consist probably of dust particles and ice (of water, ammonia and methane), while the larger rocks consist of rocky minerals covered with ice.

The ring system originated most probably together with the formation of Saturn. The structured form is caused by resonance phenomena between the periods of some inner Saturn moons and particles in the gaps of the ring system (Fig. 4.59).

We regard a moon on an orbit with radius r_m inside the ring system (Fig. 4.60). Particles with radius $r > r_m$ run slower than the moon, those with $r < r_m$ run faster. Due to the gravitational interaction between moon and particle, those with $r > r_m$ are accelerated, i.e. their angular momentum increases while it decreases for particles with $r < r_m$. Particles with $r > r_m$ are therefore pushed to larger radii, while those with $r < r_m$ are driven to smaller radii. A gap of the ring particle density is created at the moon radius $r = r_m$ (Fig. 4.60a).

There is a second mechanism, which is based on resonance phenomena: For resonances between the periods of moon and ring particles, both bodies approach a minimum mutual distance always at the same location. This causes an escalation of the perturbations and an instability of the orbit of such resonance particles. This is illustrated in Fig. 4.60b, which shows a section of the ring system and the orbits of some Saturn moons.

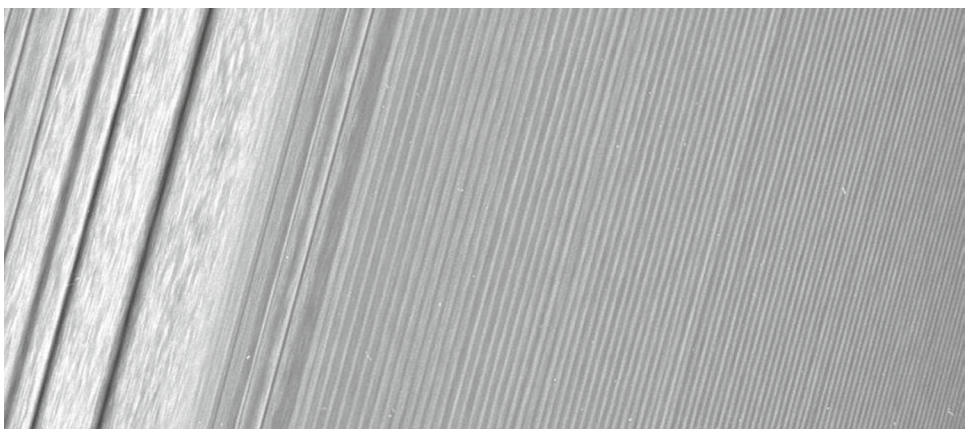


Fig. 4.58 More detailed of Saturn's ring system by the Cassini space probe (NASA)

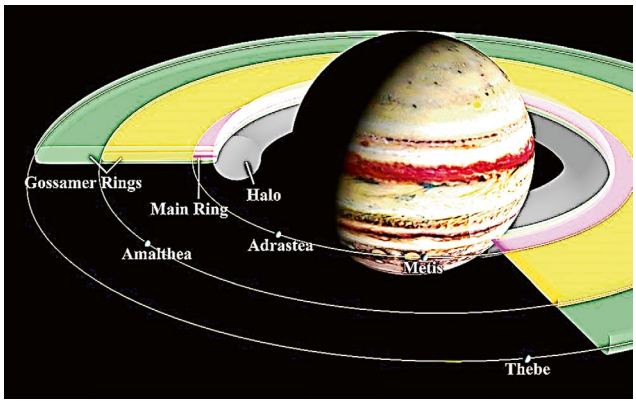


Fig. 4.59 The ring system with some moons of Saturn

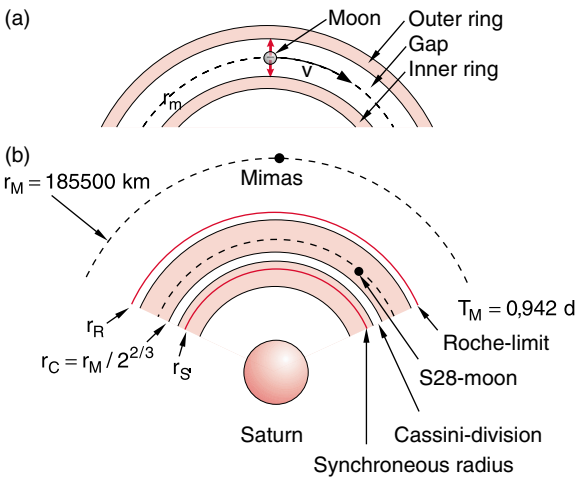


Fig. 4.60 **a** Explanation of the gaps in the ring system with the perturbing moon. **b** Synchronous radius r_s and Roche limit r_R

There are two values of the particle orbit radius of special significance: These are the **synchronous radius r_s** where the period of the particle is equal to the rotational period of Saturn, and the **Roche-limit r_R** (Fig. 4.61). For all bodies with diameter d, which orbit at radii $r < r_R$, the difference of the gravitational forces $\Delta F = F(r) - F(r+d)$ acting by the planet onto the two ends of the body is larger than the internal gravitational force F_{int} between the mass at the two ends. For example, if the body consisting of two particles with mass m bound by their mutual gravitational force, orbits around a planet with mass M inside the Roche limit $r_R = d \cdot (2M/m)^{1/3}$, the body will dissociate into two particles if $r < r_R$, i.e. $\Delta F > F_{int}$.

The inner ring system from the A to the F ring lies inside the Roche limit. Therefore, ring particles can be only stable if they are bound by forces stronger than the gravitational force, e.g. van der Waals-cohesive (chemical) forces or electro-magnetic forces. This restricts the maximum size

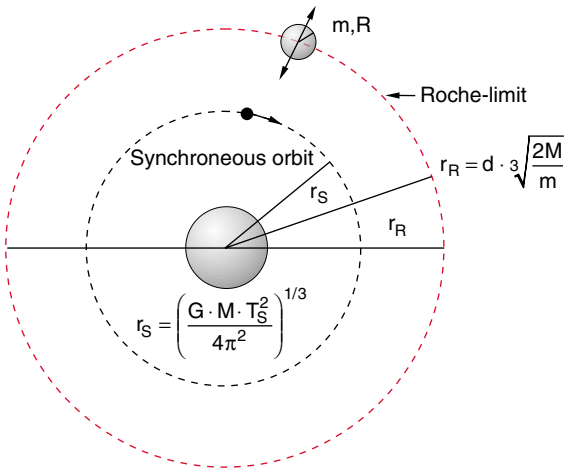


Fig. 4.61 Explanation of the Roche limit

of the particles. Larger bodies might have been formed in the beginning but were later on disrupted due to the Roche instability and form now the ring system with many small particles.

The Cassini-gap representing a distinct void in the ring system is caused by the 1:2 resonance with the Saturn moon Mimas. Particles within the gap would have a period T which is just ½ of the Mimas period. They would have been thrown out of their orbit by the periodic gravitational perturbation.

Also, the other gaps in the ring system can be explained by resonances with one of the many Saturn moons. In order to explain the stability of the existing ring system, a *Shepherd model* has been proposed (Fig. 4.62). An inner moon (1980 S26 Pandora) and an outer moon (1980 S 27 Prometheus) *called shepherd moons* tend particles in the F-ring at $r = 140.600$ km. The velocity of the inner moon at $r = 139.350$ km is slightly larger than that of the ring particles and accelerates them to larger values of r. The outer moon at $r = 141.700$ km has a lower velocity than the ring particles and decelerates them. This brings them to smaller

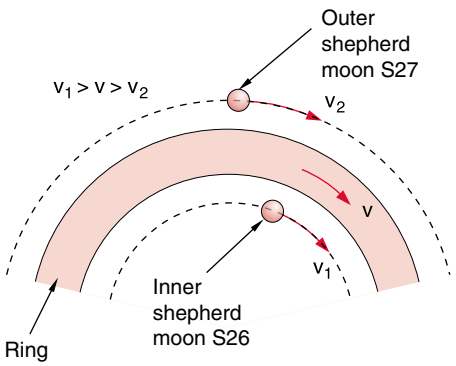


Fig. 4.62 Shepherd dog model for the explanation of the stability of a partial ring

radii r. The two shepherd moons therefore keep the ring particles within restricted r-ranges and act like sheep dogs who keep their sheep flock together.

4.4.3 The Outermost Planets

The outermost planets **Uranus** and **Neptune** have nearly equal diameters of about 50.000 km = 4.3 d_E (Fig. 4.63). Their atmospheres consist of hydrogen H_2, Helium He and Methane CH_4. They belong to the group of the 4 **Giant planets**, (also called Jupiter-like planets). However, there are certain differences to Jupiter. For example, the mean density of Uranus ($\overline{\varrho} = 1.27\,\mathrm{kg/dm^3}$) is smaller, that of Neptune ($\overline{\varrho} = 1.66\,\mathrm{kg/dm^3}$) is larger than the mean density of Jupiter ($\overline{\varrho} = 1.33\,\mathrm{kg/dm^3}$). Neptune must therefore contain heavier elements. The by far major share of the mass are for all four Giant planets Hydrogen, while Oxygen, Nitrogen, Carbon, Silicon and Iron contribute a minor share. According to new planetary models, the two planets contain a solid core consisting of silicates and metals, while the mantle is composed of methane, Ammonia and Water Vapor. The outer layers of the mantle contain mainly hydrogen and Helium, compressed by the gravitational pressure. They merge continuously into the planetary atmosphere.

Uranus is the only planet in the solar system with a rotation axis that lies nearly in its orbital plane. It is inclined against the normal of the orbital plane by 98°, causing a retrograde rotation of *Uranus*. Its magnetic field axis is tilted by 60° against the rotation axis. Its sidereal revolution period is 84 y, its synodical period is with $T_{syn} = 369.6$ d $\approx$ 1 y much smaller due to the retrograde revolution of *Uranus* [15].

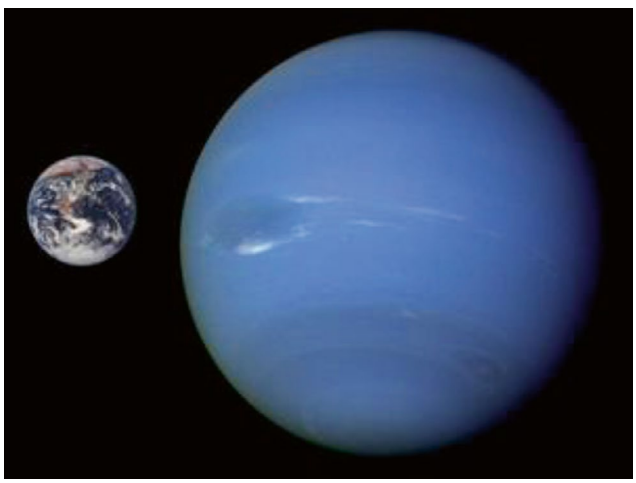


Fig. 4.63 Comparison of the sizes of Earth and Neptune (https://de.wikipedia.org/wiki/Neptun_Planet)

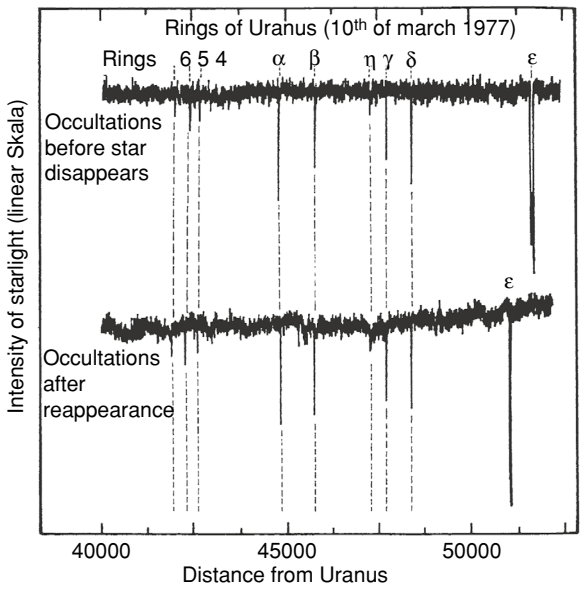


Fig. 4.64 Time variation of the radiation from the star SAO 58687 during the occultation by the Uranus rings. (From: J. L. Eliot, Ann. Rev. Of Astronomy and Astrophysics 17, 445 (1979))

Uranus has, like the other three Giant planets, a ring system that was discovered 1977 by Earth-bound observation, when the light of the star SAO158687 just behind *Uranus* was absorbed the planet and its ring system (Fig. 4.64). The transmitted light of the star detected on Earth shows incursions as a function of the observation time, when the *Uranus* passed by in front of the star. Since the velocity of *Uranus* is known, the time intervals between the incursions allow the determination of the ring radii. Some years later 1986 the ring system was also measured by the space probe Voyager 2 (Fig. 4.65).

Up to now 27 moons of *Uranus* have been discovered. The four largest are *Ariel, Umbriel, Titania* and *Oberon.* Their names are taken from theatre plays of Shakespeare.

The outermost planet of our solar system is **Neptune**. Its orbit is nearly a circle ($\varepsilon = 0.0086$, $r = 4.5 \cdot 10^9$ km = 30 AU). With a mean diameter of 49.260 km, a volume $V = 58\ V_E$ and a mass $m = 1.03 \cdot 10^{26}\,\mathrm{kg} = 17.5\ m_E$ it is the fourth largest planet in the solar system (Fig. 4.63). Its sidereal revolution period is T = 164.79 y. Its rotation axis is tilted against the normal to the orbital plane by 28°, similar to Saturn.

Neptune has been discovered in 1846 by the Germen Astronomer *Johann Gottfried Galle* some days after he had got its position from Urbain Le Vernier, who had calculated it from observed perturbations of the Uranus orbit. *Neptune* is the only planet, which has been found based on an accurate calculation.

The structure of Neptune is schematically depicted in Fig. 4.66.

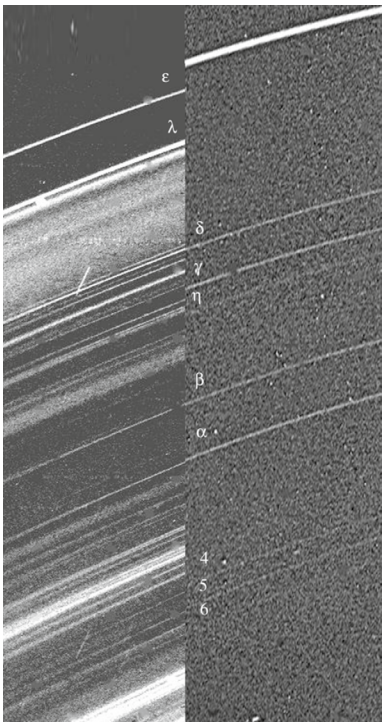


Fig. 4.65 Ring system of Uranus, photographed by Voyager 2 in forward scattered (left) and backscattered light (right) (NASA)

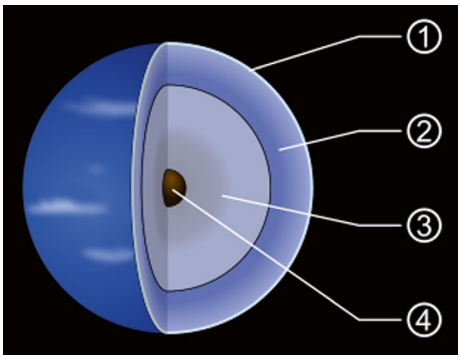


Fig. 4.66 The structure of Neptune: 1. Outer cloud layer 2. Atmosphere (H_2, He, CH_4) 3. Mantle (water, ammonia, methane) 4. Kernel (rock, ice) (https://de.wikipedia.org/wiki/Neptun_Planet)

Heavy storms have been discovered in the atmosphere of Neptune with wind velocities up to 1600 km/h.

Neptune has a ring system like the other three Giant planets. Up to now 14 moons have been discovered. Most of them are small bodies with diameters between 50 and 200 km. The three biggest are Triton (d = 2700 km), Proteus (d = 420 km) and Nereid (d = 340 km) (Fig. 4.67). Triton has about half of the mass of the Earth moon and orbits on a circle (r = 354.800 km) (Fig. 4.67). It has a thin atmosphere, which consists mainly of Nitrogen. The orbit of Nereid around Neptune is highly eccentric with a perihelion distance of 1.4 million km and an aphelion distance of 9.7 million km (Fig. 4.67).

There are altogether 181 moons in our solar system, from which 173 orbiting the 8 planets and 8 circulating around dwarf planets [17, 18].

More information about the moons in our solar system can be found in the web, e.g. in various Wikipedia articles.

4.5 Small Celestial Bodies in the Solar System

Besides the eight planets described in the foregoing section, there is a multitude of small celestial bodies in our solar system, such as the dwarf planets, planetoids, asteroids, comets, interplanetary particles (dust and gas), including particles which reach the Earth atmosphere, such as meteorites, shooting stars and particles emitted from the Sun, causing the fascinating northern light (aurora borealis).

The investigation of these bodies brings a lot of information about the formation and early history of our solar system, furthermore on the stability of the system, resonances and their influence on the structure and configuration of this many-body-system. We will therefore discuss these small bodies in the following section.

4.5.1 Dwarf Planets

Dwarf planets are celestial bodies which obey the following conditions:

- They orbit around the Sun but are not moons of planets.
- They have a sufficiently large mass to obtain hydrostatic equilibrium causing a nearly spherical shape.
- They have their surrounding not completely cleaned from small particles.

In this last point, they differ from planets

The second point allows the estimation of their minimum size.

In order to reach a spherical shape, the gravitational force must be stronger than the adhesive force between the particles.

We regard a spherical body with mass M and radius R cut into two halves. The gravitational force acting onto the intersection is

$$F_G = G \cdot (M/2)^2/\left(r_c^2\right), \tag{4.35}$$

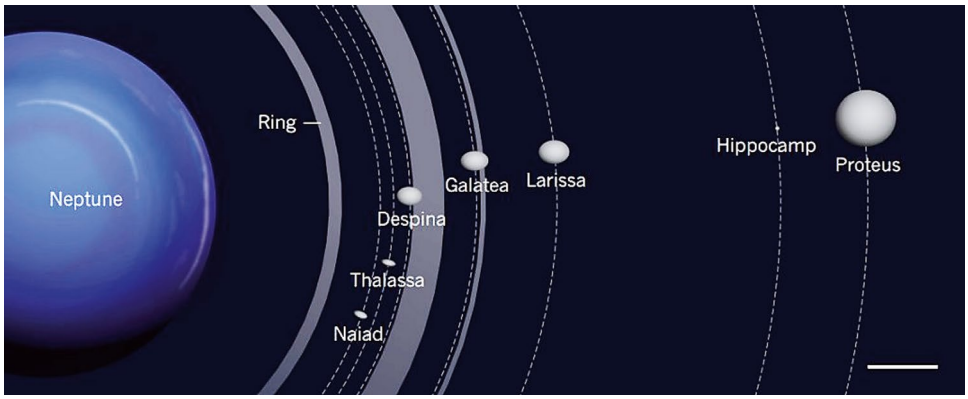


Fig. 4.67 The Moons of Neptune (https://www.bing.com/images/search?q=moons+of+neptune)

where $r_c = \left(\frac{3}{4}\right)R$ is the distance between the two centers of gravity in the two half-spheres. This gravitational force should be larger than the adhesive force

$$F_a = \pi R^2 \cdot Z \tag{4.36}$$

where Z is the adhesive force per m^2.

We then obtain for the condition $F_G > F_a$ with the density $\varrho = M/\left(\left(\frac{4}{3}\right)\pi R^3\right)$ the minimum radius R_{min}

$$R > (9/16\,\varrho) \cdot (Z/G \cdot \pi)^{1/2}. \tag{4.37}$$

Example

With $\varrho = 2.3$ kg/dm^3 and Z = 1 N/m^2 we obtain for the minimum size of a dwarf planet: $R_{min} = 600$ km. ◀

The size of the dwarf planets is estimated from their brightness, which is determined by the product of the incident solar radiation and the albedo (ratio of incident to emitted radiation power). Since such measurements have relative large uncertainties, the size of the dwarf planets can be only measured within large error bars.

Up to now, five objects could be classified as dwarf planets, these are: Pluto, Eris, Makemake, Haumea and Ceres. It can be, however, expected, that much more objects in the Kuiper belt can be classified into the category of dwarf planets [19]. There are already about 20 candidates under investigation, which might fulfil the criteria for dwarf planets.

We will shortly present the most important data for the up to now observed and measured five dwarf planets (Fig. 4.68 and Table 4.7).

Ceres is the closest dwarf planet to the Sun. It is located in the *asteroid belt* between Mars and Jupiter. It has been already discovered in 1801 by *Giuseppi Piazzi*. With its mass $m = 9.35 \cdot 10^{20}$ kg $= (1/6360)\ m_E$, it is the smallest of the five dwarf planets.

Pluto was discovered in the Kuiper Belt only 1930 by the *Lowel observatory* in Arizona at a careful sky survey after predictions of its approximate position deduced from observed perturbations of the Neptune orbit. The inclination angle of its strongly eccentric orbit ($\varepsilon = 0.2488$) against the elliptical plane $i = 17.16°$ is much larger than that of all other planets. This fact points already to its outsider role in the solar system.

Its mass $m = 1.25 \cdot 10^{22}$ kg amounts only to 0.22% of the Earth mass m_E. Its diameter D = 2390 km is less than 1/5 of the Earth diameter.

Its distance from the Sun varies between 29 and 49 AU. It moves partly inside and partly outside of the Neptune orbit (Fig. 4.69). Its sidereal period is $T_{sid} = 247.68$ a and its rotational period is $T_{rot} = -6\,d\,9\,h\,17$ min. The minus sign indicates that Pluto rotates (like Venus) retrograde.

Up to now, five moons of Pluto (*Charon, Nix, Hydra, Kerberos* and *Styx*) have been discovered.

Eris is the most massive and second largest of the known dwarf planets with a mass $m = 0.0025\ m_E$ and an equator diameter D = 2400 km. It moves in the Kuiper Belt between distances from the Sun $r_{max} = 97.5$ AU in the aphelion and 37.8 AU in the perihelion of its strongly eccentric elliptical orbit ($\varepsilon = 0.43$). Its orbital period is $T_{sid} = 559.07$ a and its rotational period is $T_{rot} > 3$ d.

Makemake was discovered 2005 at the Palomar Observatory. Its name is taken from the Creation God of the Easter Island. With its diameter of D = 1477 km it is the largest object in the Kuiper Belt. Its orbital plane is inclined against the ecliptic by the angle $i = 29°$. The eccentric orbit ($\varepsilon = 0.156$) has the major axis a = 45.6 AU and its orbital period is $T_{sid} = 307.5$ a. The perihelion distance from the Sun is 38.36 AU and its aphelion distance 52.76 AU. With the apparent brightness of 17 mag, it is after Pluto the second brightest object in the Kuiper Belt. Its rotational period is $T_{rot} = 7.77$ h. It has one moon (Fig. 4.70).

Haumea (name of the Hawaian goddess of fertility) has about the same size as *Pluto*, deviates, however, strongly from a spherical shape, but has the shape of strongly stretched ellipsoid with diameters of 2200 and 1100 km resp. Its rotational period is with $T_{rot} = 3.9$ h very short. It moves on a strongly eccentric orbit ($\varepsilon = 0.189$) which

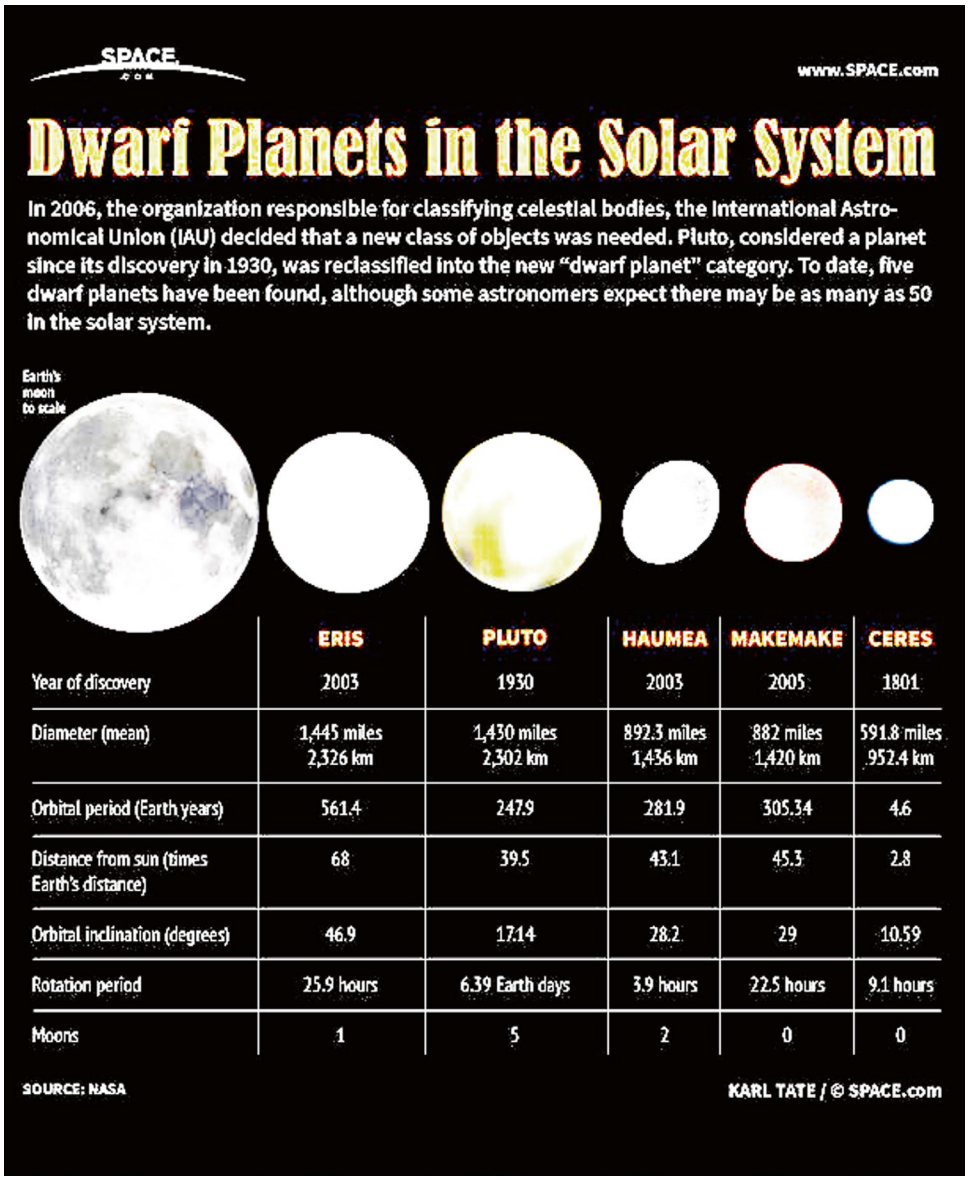


Fig. 4.68 Dwarf planets in our solar system (Karl Tate Space com)

Table 4.7 Dwarf planets in our solar system

Name	Eris	Pluto	Makemake	Haumea	Ceres
Mass in units of the Earth moon mass	0.2	0.178	0.08	0.053	0.021
Large semi-axis of the orbit in AU	68.1	39.5	45.7	43.3	2.76
Numerical eccentricity	0.43	0.247	0.156	0.189	0.078
Sidereal revolution period in years	562.5	247.68	308.5	285.3	4.601
Sidereal rotation period in days	> 0.3	−6.39	?	0.16	0.378
Number of moons	1	3	0	2	0

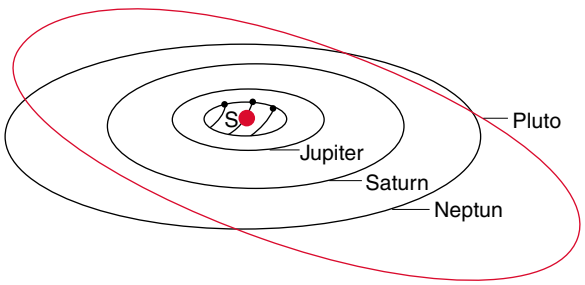


Fig. 4.69 The orbital plane of the dwarf planet Pluto, inclined against the orbital plane of the outer planets

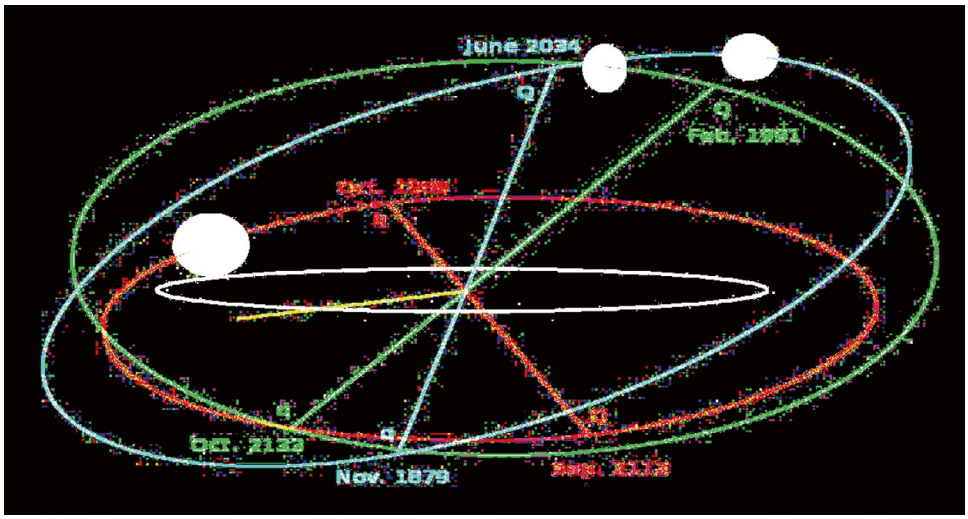


Fig. 4.70 Orbits of Makemake (blue), Pluto (red), Haumea (green) and Neptune (grey). The orbit of Neptune is not too scale. It should only illustrate the inclination of the dwarf planets' orbital planes (https://commons.wikimedia.org/w/index.php?curid=740911)

is inclined against the ecliptic by i = 28°. Its distances to the Sun vary between 35 and 51 AU. The orbital period is $T_{sid} = 285$ a. Up to now two moons have been discovered.

Further candidates for dwarf planets are.

Quoar has been discovered in 2002. Its name was chosen from the creation myth of the north-American Tongva-Indians. Its diameter is D = 900–1200 km and its mass is 1–2 10^{21} kg ($3 \cdot 10^{-4}$ m_E). Its nearly circular orbital path ($\varepsilon = 0.036$) has the inclination angle $i = 8°$ against the ecliptic and its major axis is $a = 43.6$ AU. The orbital period is $T_{sid} = 288.1$ a.

Sedna (Inuit Goddess of the Sea) was discovered in 2003 at the Mount Palomar Observatory. With its estimated diameter D = 1600–1700 km it is the second largest object in the Kuiper Belt. Its very eccentric elliptical orbit ($\varepsilon = 0.841$) has distances from the Sun between (13 and 130) $\cdot 10^9$ km (76–884 AU). The extremely large eccentricity is probably due to perturbations by Neptune.

Orcus (name of Roman God of Netherworld) was discovered in 2004 at the Astronomical Observatory of CALTEC. Its diameter is D = 1600–1800 km.

4.5.2 Planetoids

The range between the orbits of Mars and Jupiter (2–3.4 AU) is populated by many small celestial bodies with diameters between 1 and 1000 km. They cannot be observed with the naked eye, but with telescopes more than 3000 of such Planetoids (also called Asteroids) have been discovered. Their total number is estimated to be about 1 million. However, the total mass of all planetoids just reaches about 5% of that of the Earth moon.

Plotting the mean distances a_i of the planets from the Sun, according to the Titus-Bode empirical rule

$$a_i = 0.1 \cdot (4 + 3 \cdot 2^n)\ \text{AU} \tag{4.38}$$

against the integer n, the results of Table 4.8 are obtained. The table shows that there is no planet for n = 3 ($a = 2.8$ AU). This just corresponds to the mean distance a of the planetoid belt. Although the range Δa of the planetoid belt is small, its volume $4\pi a^2 \cdot \Delta a$ is larger than the volume $V = 4/3\pi a^3 (a < a_{Mars})$ of the sphere inside the Mars orbit.

The planetoids have found increasing interest since it became obvious, that a large fraction of the number of meteorites, falling onto the Earth, originates from the planetoid belt. In order to understand the origin of the asteroid belt, we have to collect all known facts. Plotting the numbers of all planetoids with diameters D > 80 km against their distances a, one obtains the diagram in Fig. 4.72. It is immediately apparent that there are gaps in the distribution

Table 4.8 Calculated planetary distances from the Sun based on the Titus-Bode relation and compared with the observed large semi-axis of the planets

Planet	n	Distance/AU	
		Calculated	Observed
Mercury	$-\infty$	0.4	0.4
Venus	0	0.7	0.7
Earth	1	1.0	1.0
Mars	2	1.6	1.5
	3	2.8	–
Jupiter	4	5.2	5.2
Saturn	5	10.0	9.2
Uranus	6	19.6	19.2
Neptune	7	38.8	30.1
Pluto	8	77.2	39.5

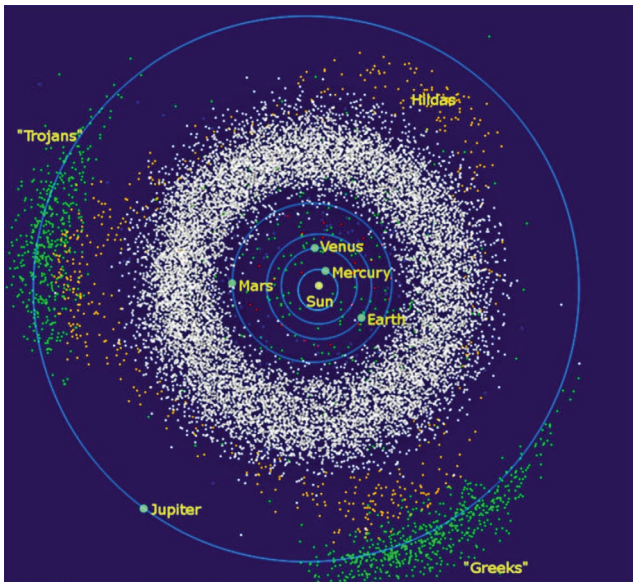


Fig. 4.71 The Planetoid belt (https://www.bing.com/images/search)

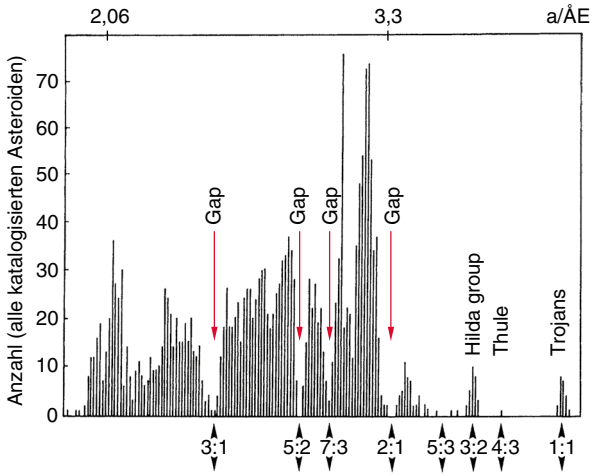


Fig. 4.72 Number of planetoids with radii R > 40 km plotted against their distance from the Sun. The arrows indicate resonances of the revolution times $T_{plantoid}/T_{jupiter}$ (after J. K Beatty: Die Sonne und ihre Planeten. Physik Verlag Weinheim 1985)

N(a) where for certain values of a no planetoids are found. How can this be explained?

When the ratio $T_{plan}/T_{jupiter}$ of the orbital periods of a planetoid and of Jupiter are calculated according to the 3rd law of Kepler, one obtains for the gaps the ratio n_1/n_2 with small integer numbers n_1 and n_2.

These special ratios lead to stronger perturbations due to the gravitational interaction between the massive Jupiter and the small planetoid resulting in an instability of the planetoid orbit, which can be seen as follows.

For the ratios $T_{plan}/T_{jup} = n_1/n_2$ of small integers the planetoid comes after n_2 revolutions ($t_1 = n_2 \cdot T_{plan}$) at the same location relative to Jupiter ($t_2 = n_1 \cdot T_{jup} = t_1$). Therefore, the perturbations add up until they have thrown the planetoid out of its orbit.

At these ratios, therefore, no planetoids can be found any longer and gaps appear in the distribution N(a) of the number of planetoids. These gaps are called *Kirkwood Gaps* (Fig. 4.72). These resonance-based perturbations can change the orbit of a planetoid in such a way that it may collide with other planetoids. Such collisions can even make the planetoid to leave its orbital plane and may eventually cause collisions with other planets. Such planetoids passing through the Earth atmosphere appear as bright meteorites.

Since the eccentricities of the different planetoids can differ, a planetoid may move on an orbit that cross the orbits of other planetoids, this increases the probabilities of collisions. Because of the high impact energy such collision can shatter the collision partners. The fragments can be pushed to orbits that cross the Earth orbit (Fig. 4.73). If they pass through the Earth atmosphere they are decelerated due to friction, heated up to very high temperatures and appear as bright meteorites. If they have sufficient mass they can reach the Earth surface and create impact craters.

There are two interesting exceptions from the instabilities of planetoid orbits:

1. The *Trojans* and the *Greek* ($T_{plan}/T_{Jup} = 1:1$) and
2. The *Hilda Group* ($T_{plan}/T_{jup} = 1:2$).

The Trojans and the Greek represent two groups of about 100 asteroids (Fig. 4.71) which have the same period T_{sid} as Jupiter and must have therefore the same distance to the Earth (Fig. 4.75a). They are located in *Lagrangian points* (also called *Libration points*). Due to the superposition of the gravitational fields of the Sun and a planet, there are five Lagrangian points where the potential energy has minima in these points (Fig. 4.74) and therefore the total force on a body in these points is zero.

Joseph Louis Lagrange had already 1772 recognized that at certain constellations of three bodies the interaction between these three bodies results in stable configurations with minima of the total potential energy. The total force on bodies in these Lagrangian points as gradient of the potential energy is then zero. For these constellations, the three body problem can be exactly solved. We will illustrate this with Fig. 4.75a (The corresponding calculation can be found in Problem 4.13).

When two large masses (in this case Sun and Jupiter) move around their common center of mass, there are altogether five Lagrangian points: The points L_1, L_2 and L_3 are saddle-points of the potential energy, while L_4 and L_5 represent local minima of the potential energy, which are located at the corners of the equilateral triangles $M_1M_2L_4$ and

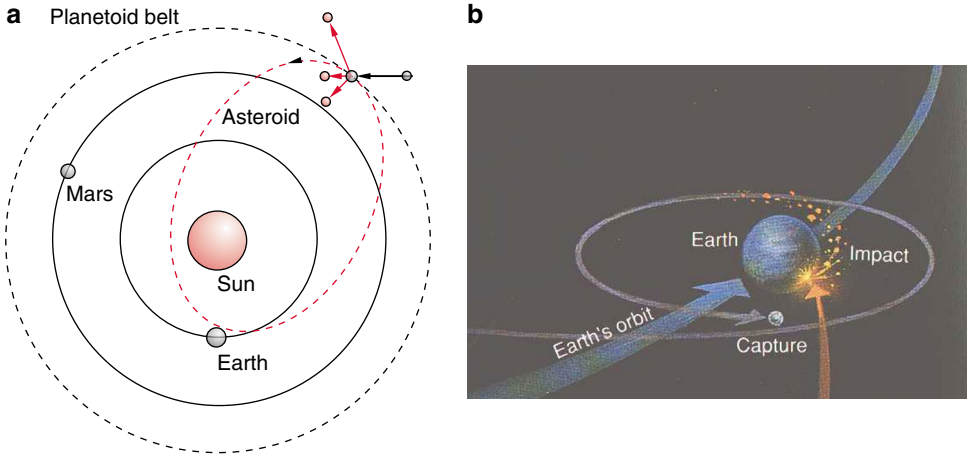


Fig. 4.73 **a** Orbits of Asteroids close to the Earth orbit caused by collisions between planetoids. **b** A planetoid from the asteroid belt can be thrown by collisions into an orbit crashing with the Earth (Wikipedia)

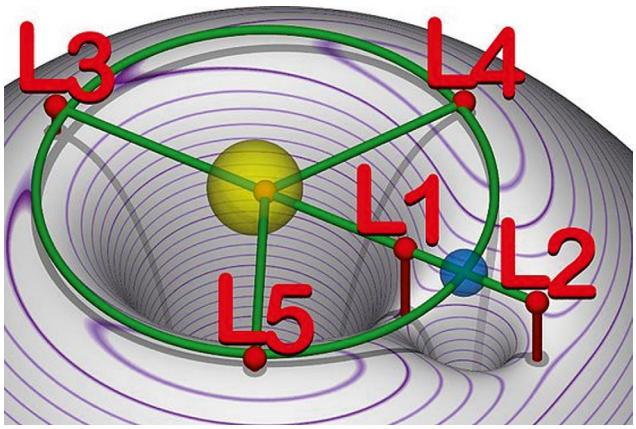


Fig. 4.74 Potential energy minima in the Lagrangian points L_1–L_5 (Wikipedia)

$M_1M_2L_5$. In these two points L_4 and L_5, a large number of planetoids are assembled. Their names were chosen as the names of heroes in Homer's Ilias. In L_4 are the defenders of Troja (Patroclus, Priamos, Hector etc.), in L_5 the attacking Greeks (Achill, Agamemnon etc.).

The planetoids captured in these Lagrangian points can perform motions around the local minima, induced by perturbations by other close by planetoids. These motions are called **Librations** (Fig. 4.75b). Therefore, these local minima are also called **Libration points**. As long as the perturbations are sufficiently small, the planetoids do not leave their potential energy trough but perform motions with limited amplitude around their Libration point.

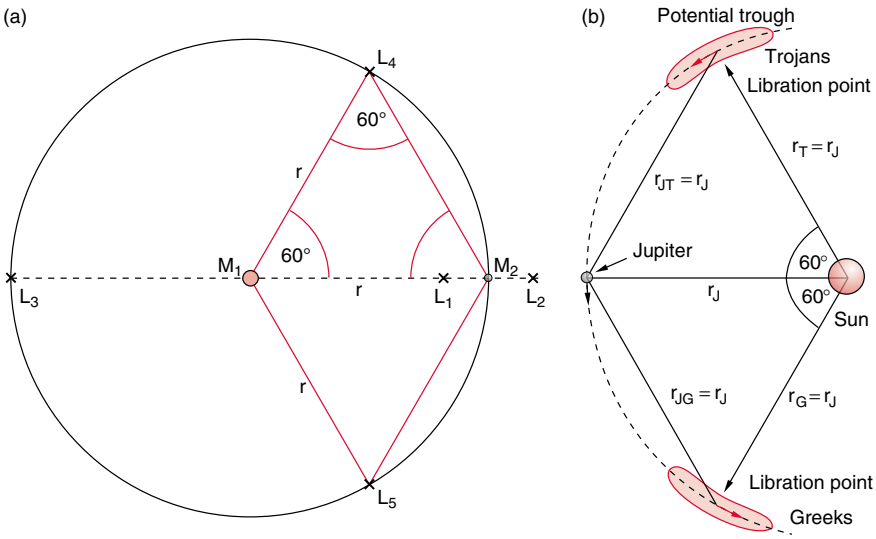


Fig. 4.75 **a** Lagrange points L_i caused by the gravitational interaction between two heavy masses M_1 and M_2 and a light body m in L_1 or L_2. **b** Motion of the Trojans and the Greek asteroids in the potential trough around their libration points

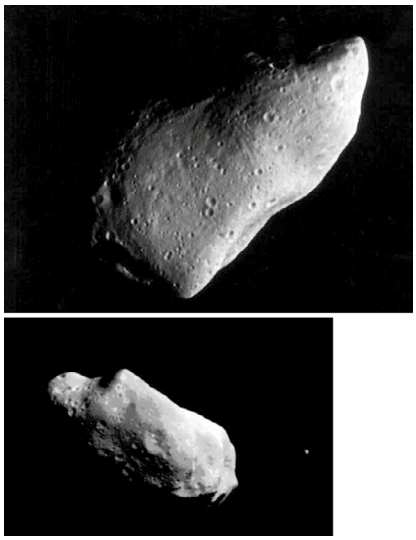


Fig. 4.76 Planetoid pair Ida and Gaspra (NASA)

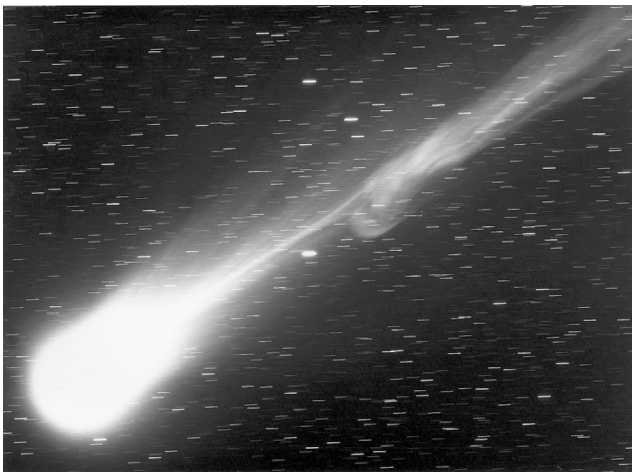


Fig. 4.77 Comet Hayakutake (photographed by E. Kolmhofer, H. Raab; Johannes-Kepler-Observatory, Linz)

Fig. 4.78 The comet Hale-Bopp (photographed by H. Perret, SAGA, Uni-Kaiserslautern)

Different models for the size and shape of the planetoids have been developed, depending on the experimental approach.

Regarding the radiation intensity I(t) reflected by the larger planetoids, one often observes a periodic variation. This was initially attributed to the rotation of elongated bodies. Later more detailed investigations showed, however, that at least part of this periodic reflected intensity is due to double-body- systems, orbiting around their common center of mass. One example is the planetoid pair Ida and Gaspra (Fig. 4.76). Generally, the shape of larger planetoids is approaching a spherical form, while smaller planetoids can have very irregular shapes.

Most of the planetoids are extremely dark, i.e. their albedo is with A = 0.03–0.04 very small. This indicates that they consist of carboniferous chondrites. According to our present knowledge planetoids are cold celestial bodies without an atmosphere.

There are two models about the origin of the planetoids: They could have been either formed by fragmentation of larger bodies or they could have been generated already at earlier times when the planets of the solar system had been formed by agglomeration of small dust and rock particles, like the formation of the planets.

4.5.3 Comets

The phenomenon of a bright comet with a tail on the sky has fascinated people since early days. It was often regarded as herald of catastrophic events. The name comes from the Greek name (*no μητηζ* = longhaired, feathered). A particular impressive recent example is the comet Hyakutake, which passed the Earth 1996 so close, that it could be observed in detail by the naked eye (Fig. 4.77).

One of the brightest comets in the modern period is *Hale Bopp*, which could be seen 1996/1997 with the naked eye and which belonged for several months to the brightest objects on the sky (Fig. 4.78).

Each year about 10–20 new comets are discovered. Most of them have, however, such a low intensity that they need telescopes with high light-gathering power in order to observe them. Up to now more than 2700 comets have been registered.

One distinguishes between periodic and aperiodic comets. The periodic comets are subdivided into short and long periodic comets. The orbits of long periodic comets are highly elongated elliptical orbits with major axes reaching from a = 40.000 AU up to a = 150.000 AU.

Fig. 4.79 Photo of the comet Halley (taken by Dr. K. Birkle on the 09.01.1986 with the Schmidt telescope of the CALAR ALTO observatory of the MPI for Astronomy Heidelberg)

The short periodic comets have period s T between T = 2–200 years and move on elliptical orbits with major axes a between a = 2–30 AU.

A famous example for a short period comet is the comet Halley (Fig. 4.79) (period T = 76 y, major axis $a = 18$ AU, eccentricity $\varepsilon = 0.967$). The perihel has a distance of 0.6 AU, its aphelion lies close to the Neptune orbit. It, therefore, comes closer to the Sun than the planet Venus. This Haley comet is assumed to be the celestial object which guided the astronomers at Christ's birth to Betlehem.

A comet consists of a *solid core* with diameters of some km (Fig. 4.80), a *halo* (gaseous envelope of atoms, molecules and ions with a diameter up to 100.000 km and a comet tail, which develops when the comet approaches the Sun and which can have a length of many million km (Fig. 4.79).

When the comet approaches the Sun at distances <3 AU, the Sun radiation causes the evaporation of a small part of the comet core. At the dark spots on the surface, where the absorption of the Sun radiation is higher, jets of evaporation material shot out of the surface, as can be seen on the photo of the core of the Halley comet (Fig. 4.80) taken by the Giotto space probe passing the comet at a close distance from the core. This photo was the first picture of a comet core, which has a size 16 × 8 × 8 km and the shape of a big potato.

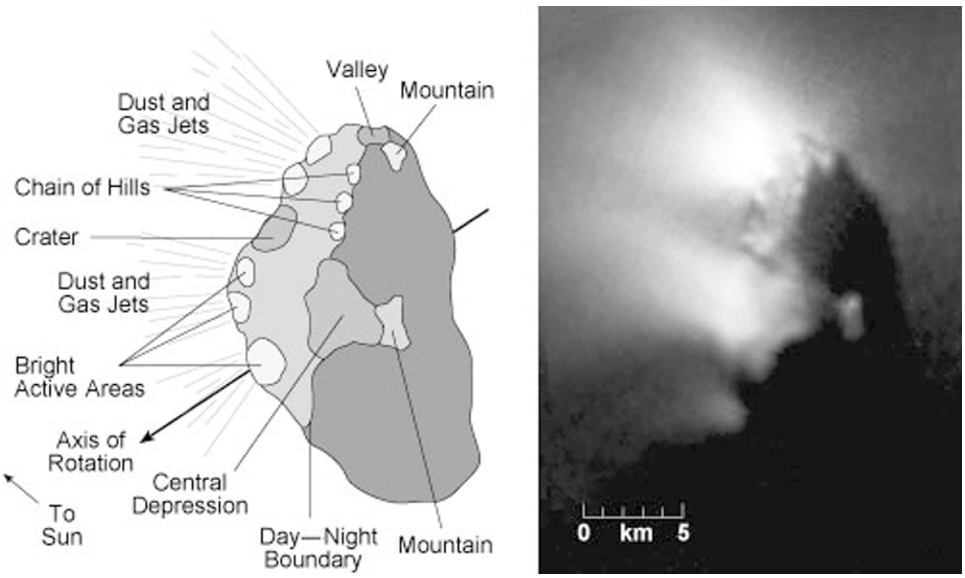
Up to now, about 50 comets have been found, which move on hyperbolic orbits, i.e. they pass only once through the solar system [21, 22].

The present state of research is, after measurements of more than 600 orbits of comets, that all comets originally had been long period comets. Their origin lies in the Oort's cloud at a distance from the Sun ranging from 2000 to 5000 AU at its inner edge until 100.000 AU at its outer edge.

Perturbations of their orbits while they passed close to a planet have changed their orbits. If they have the same sense of revolution as the planet, they are accelerated when passing the planet (Swing by). They then can get enough kinetic energy to leave the solar system. If they have the opposite sense of revolution, they are decelerated and can be transferred to short period elliptical paths and come closer to the Sun (see Problem 4.9).

A spectacular example of the perturbation of a comet orbit by a planet is the impact of the comet Shoemaker-Levy onto Jupiter. The comet passed Jupiter 1992 so close that it came below the Roche limit (see Fig. 4.61). The result was a fragmentation of the comet into 21 fragments which were distributed over several million km along the comet path (Fig. 4.81). Between June 16 and 24 in 1994, they entered the atmosphere of Jupiter and were decelerated by friction and crashed into Jupiter. This was the first time that a collision between two celestial bodies could be observed from the Earth.

Fig. 4.80 The Nucleus of Comet Halley (MPI for Astronomy Lindau, Harz Germany and ESA)



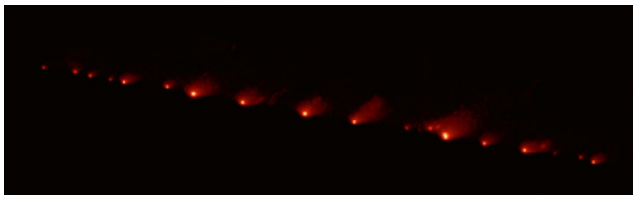


Fig. 4.81 The comet Shoemaker-Levy was disrupted into 21 fragments after it surpassed the Roche limit of the Jupiter. The fragments crashed later on into Jupiter (Comet Shoemaker-Levy 9-Wikipedia)

All observations indicate that the comet core is like a big dirty snowball, i.e. an irregularly formed body of water ice, interspersed with dust, carbon- and silicon-containing minerals and frozen CO_2 and NH_3. It is often covered with a dark crust.

The ESA space probe Rosetta orbited around the comet Tschurjumv-Gerassimenko, after a complicated way through the solar system [23]. It could send a lander onto the core of the comet Tschurjumov and took brilliant photos of its surface (Fig. 4.82). The size of the core is 4×3.5 km and its mass is m = 1013 kg. It rotates around its center of mass with a period T = 12.7 h and it revolves in 10.7 y on a strongly eccentric orbit ($a = 3.46$ AU, $\varepsilon = 0.64$) around the Sun. The apparent very irregular surface and the strange shape of the core had been very surprising for the scientists.

The material evaporating with a velocity of 0.5 km/s from the solid comet core, forms the gaseous spherical halo of the comet (Fig. 4.83). The density of the coma is about 10^4–10^6 atoms or molecules per cm^3. The molecules are dissociated by the UV radiation of the Sun, leaving atoms, ions and radicals as main material of the coma. The ions are deflected in the magnetic field of the extended Sun magneto-sphere and driven away from the comet by the solar wind (particles emitted from the Sun). This generates the long comet tail which can extend up to 10^8 km, but has a very low density of 10–100 mol/cm^3. It would therefore not harm mankind if the tail of a comet would pass through

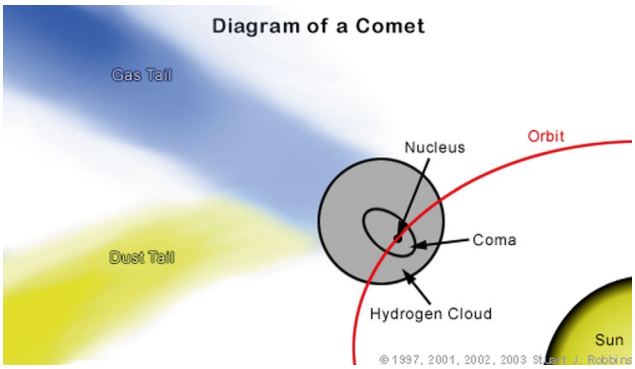


Fig. 4.83 Schematic drawing of the composition of a comet (not to scale!) (Wikipedia)

the Earth, as had occurred 1910 when the tail of the Haley comet had contact with the Earth.

Observations show that the tail often consists of several components: One consisting of neutral particles (dust tail) and one of ionized particles. While the neutral part is deflected by the radiation pressure away from the Sun, the ionized part is deflected by the magnetic field.

This is illustrated by Fig. 4.84, which shows the photograph of Comet West, one of the largest comets of all-time, taken by the amateur astronomer John Laborde. The thin blue tail is the ion tail, while the broad white tail is formed by dust particles.

The comet loses due to this gas evaporation about 0.1–1% of its mass per revolution around the Sun, until its ice mantle has been evaporated. Then the evaporation rate decreases.

4.5.4 Meteors and Meteorites

Meteors (also called shooting stars) are short time brilliant light tracks on the night sky, caused by small interplanetary dust particles (meteorites) passing through the Earth

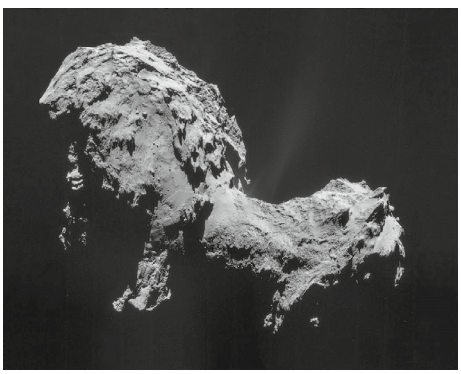


Fig. 4.82 Komet Tschurjumov-Gerassimenko. *Photo* ESA/Rosetta/NAVCAM, CC BY-SA IGO 3.0, CC BY-SA 3.0 igo

Fig. 4.84 Comet West and its ion (blue) and dust (white) tail (John Laborde https://www.googleadservices.com)

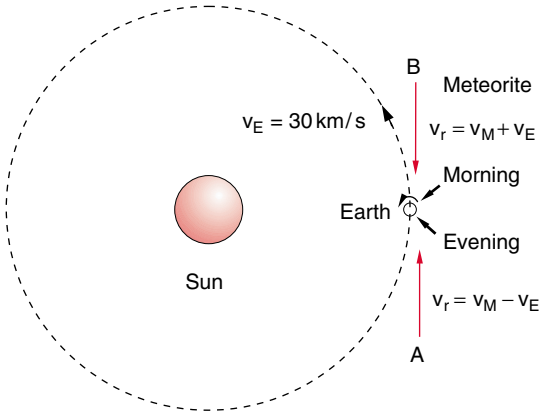


Fig. 4.85 Different relative velocities of meteorites against the Earth depending on their velocity parallel or antiparallel to the Earth velocity

atmosphere. Due to friction, they are heated up to such high temperatures that they emit visible radiation. Small particles vaporize in the atmosphere while larger bodies reach the surface of the Earth and can be found. One example is the Barringer Crater in Arizona, created by the impact of a medium sized meteorid.

The light track is generated by a channel of excited or ionized molecules which are produced by impact of the fast meteorites (dust particles or mineral lumps) with the atmospheric molecules (relative velocities up to 70 km/s). The altitude range where the meteorites can be seen is 10–100 km where the maximum light intensity of the light track appears between 80 and 100 km. A meteor is only visible for about 1 s.

The relative velocity of the meteorites relative to the Earth depends on their incidence direction (Fig. 4.85). In the beginning of the night, the observer is on the "backside" of the Earth orbiting around the Sun. Meteorites reaching the observable part of the night sky come from the direction A. Their relative velocity $\boldsymbol{v}_r = \boldsymbol{v}_M - \boldsymbol{v}_E$ is smaller than in the early morning at the end of the night where the observer sits on the front side of the Earth and observes meteorites coming from the direction B. Their relative velocity is now $\boldsymbol{v}_r = \boldsymbol{v}_M + \boldsymbol{v}_E$.

Example

For typical velocities $v_M = 40\,\text{km/s}$ of meteorites, we obtain with $v_E = 30\,\text{km/s}$ the relative velocities $v(A) = 10\,\text{km/s}$ and $v(B) = 70\,\text{km/s}$. ◀

Most of the particles penetrating into the Earth atmosphere have diameters ranging from a few μm to 1 cm and masses between milligrams to some grams. There are also meteorites with masses of several kg to many tons. Because of their

Fig. 4.86 Barringer Crater in Arizona (Wikipedia)

larger mass they do not vaporize in the atmosphere but reach the Earth surface and cause deep impact craters (Fig. 4.86).

Per year about 20.000 meteorites with masses $m > 100\,g$ fall onto the Earth surface. The total mass of all meteorites entering the atmosphere per year is $1.6 \cdot 10^7\,\text{kg/y}$.

The small meteorites appear cumulated at certain days of the year, when the Earth crosses the remnants of earlier comet tails (shooting stars showers). The increased particle density in these comet tails causes much more shooting stars per night than at other times. Examples are the *Perseide showers* (27 July–17th of August), the *Leonide flights* (11–20 November) and the *Geminide showers* (6–16 December).

The origin of the large meteorites was for a long time unknown. After many observations of the orbits of meteorites (which can be only followed over a short path length), it turned out that most of the large meteorites come from the planetoid belt between Mars and Jupiter. When a planetoid is thrown out of its orbit through perturbations by Jupiter or Mars, it has a larger probability of collisions with other planetoids. Such collisions can shatter the planetoid into many smaller fragments which then orbit around the Sun where their orbital plane can be strongly inclined against the ecliptic.

This meteorite can cross the Earth orbit and eventually enter the Earth atmosphere (Fig. 4.73), where it is slowed down due to friction and becomes so hot, that it appears as a shooting star. Since its initial velocity is generally larger than the escape velocity $v_{es} = 11.2\,\text{km/s}$ (see Vol. 1. Sect. 2.6.3), it may pass tangential through the higher atmosphere and leave the Earth again.

The friction heats the meteorite up to such high temperatures that part of its surface evaporates and a hot plasma sphere develops around the meteorite, which might even reach the surface temperature of the Sun. The large temperature gradient between surface and inner part of the meteorite may induce such large mechanical tension that

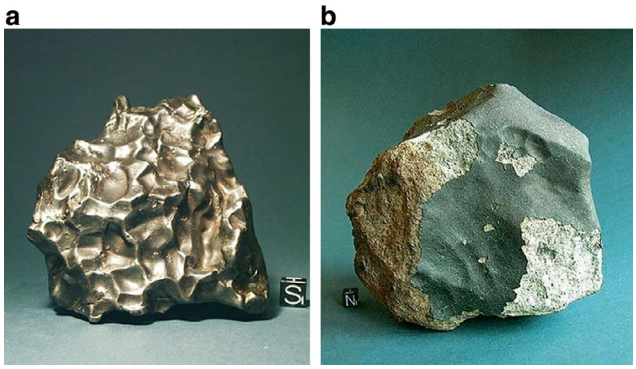


Fig. 4.87 **a** Iron-Meteorite Sikhote-Alin 1.3 kg. **b** Stone-Meteorite found in the Sahara (Wikipedia)

the meteorite bursts into many fragments, which produce a meteorite shower.

About 95% of all meteorites found on Earth consist mainly of rocky material with small admixtures of metals. About 5% contain mainly metals (e.g. 90% iron and 8–9% nickel). The mineral meteorites are divided into *Chondrites* and *Achondrites*. The Chondrites contain spherical inclusions (d = 0.1–1 cm) of Olivin and other minerals. They can be explained by melted droplets, which have been formed at collisions of planetoids.

Achondrites are essentially metal-rich mineral meteorites, which have been produced during volcanic eruptions on planets. The chemical and mineralogical analysis gives information about the conditions during the formation of the meteorite. The radioactive age determination informs about the formation time of the meteorite. The formation age is that time after a high temperature phase when the meteorite cooled down for the first time and formed a closed chemical system, which did no longer change during the whole time until it was found (Figs. 4.87 and 4.88).

The amount of fission products produced by the cosmic rays in the meteorite gives information about the time interval the meteorite had been exposed to cosmic radiation. Since this radiation can penetrate only a few meters into the meteorites, the analysis allows the determination of the time when the planetoid fragmented (see Sect. 6.8).

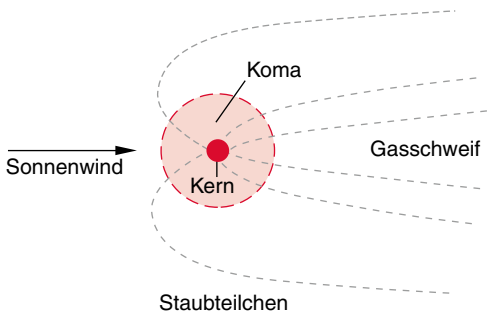


Fig. 4.88 Schematic composition of a comet

4.6 The Sun as Stationary Star

Many characteristic properties of the Sun change in the course of time so slowly that we can regard them as constant. Such constant properties and the experimental techniques for their measurements will be discussed in the following section. These constant properties are the size and the mass of the Sun, its surface temperature and the luminosity of the Sun. We will learn, how the knowledge of these basic properties, together with known laws of Physics, will lead us to the development of a realistic model of the Sun, which correctly explains all observations. This model allows furthermore the description of the interior of the Sun, although the solar neutrinos are the only source of direct information about the processes in the inner part of the Sun. All other observations, such as the solar radiation from the surface of the Sun, the magnetic effects and the periodic oscillations of the size of the Sun give indirect information, which needs the interpretation of the results based on a Sun model [24].

Besides these stationary properties, there are dynamical short-term variable characteristics, as for example the sunspots, the granulation of the Sun surface, eruptions within the Sun surface, such as flares and protuberances. They will be treated in Sect. 4.7.

In Chap. 5, it will be shown that our Sun is a quite ordinary star with medium luminosity. Since it is much closer to the Earth (8 light minutes) than the next other star (Proxima Centaury, distance 4 light years), we can study general star properties in much more detail on our Sun than on far distance stars. We can then use models obtained from observations of the structure and the development of our Sun to explain properties of other stars with comparable mass.

4.6.1 Mass, Size, Density and Luminosity of Our Sun

Measurements of the revolution period of the Earth $T = 1$ year and the major axis a of its elliptical orbit allow the determination of the mass $M_\odot$ of the Sun according to Eq. 4.13

$$\frac{T^2}{a^3} = \frac{4\pi^2 \cdot M_\odot}{G \cdot (m + M_\odot)^2}$$

Inserting numerical result, one obtains

$$M_\odot = 1.989 \cdot 10^{30}\,\text{kg}.$$

Observing the Sun through a strongly attenuation filter, it appears as perfect circle with the angular diameter of $2\alpha = 31'59''.2 = 0.009204$ rad (Fig. 4.89).

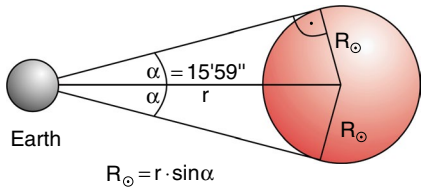


Fig. 4.89 Measurement of the Sun radius

The radius of the Sun is therefore

$$\boldsymbol{R_\odot} = 1\mathbf{AU} \cdot \sin\boldsymbol{\alpha} = 6.96 \cdot 10^8\,\mathrm{m} = 696.000\,\mathrm{km}.$$

This gives the volume of the Sun

$$V_\odot = \frac{4}{3}\pi R_\odot^3 = 1.3 \cdot 10^{27}\,\mathrm{m}^3$$

The mean density is then

$$\overline{\varrho} = M_\odot / V_\odot = 1.41\,\mathrm{kg/dm}^3,$$

This is about ¼ of the Earth density.

The luminosity $L_\odot$ of the Sun, i.e. the total radiation power emitted into the full solid angle 4π can be determined from the measured solar constant (Fig. 4.90).

The solar constant SC is the radiation power incident onto 1 m^2 of the Earth surface outside the Earth atmosphere.

$$SC = \frac{L_\odot}{4\pi r^2}\,\mathrm{W/m}^2 \quad \text{with} \quad r = 1\,\mathrm{AU}. \tag{4.39}$$

The solar constant is measured with bolometers with blackened napped surface A, which ensures that the incident radiation is completely absorbed. The radiation power absorbed per time interval dt results in a temperature increase

$$\mathrm{d}T = A \cdot \frac{SC}{c_v \cdot m} \cdot \mathrm{d}t, \tag{4.40}$$

of the bolometer with mass m, which is carefully temperature isolated against its surrounding.

Such measurements give the value of the **solar constant**

$$\boldsymbol{SC} = 1.36 \cdot 10^3\,\mathrm{W/m}^2. \tag{4.41}$$

This yields according to (4.39) the luminosity of the Sun

$$\boldsymbol{L_\odot} = 3.82 \cdot 10^{26}\,\mathrm{W}.$$

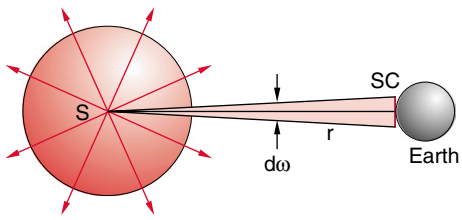


Fig. 4.90 Determination of the luminosity of the Sun by measuring the solar constant

Using the relation $E = m \cdot c^2$, the Sun loses due to its emitted radiation per s the mass

$$\frac{\mathrm{d}M_\odot}{\mathrm{d}t} = 4.2 \cdot 10^9\,\mathrm{kg/s}.$$

If we assume that the Sun during its present lifetime of about 5 Billion years ($5 \cdot 10^9 y$) had the same luminosity as today, we obtain a total energy loss of

$$\frac{\mathrm{dW}}{\mathrm{d}t} \cdot \Delta t = 3.8210^{26}\,\mathrm{W} \cdot 5 \cdot 10^9 : 365 \cdot 24 \cdot 3600\,\mathrm{Ws} = 6 \cdot 10^{43}\,\mathrm{Ws}$$

and a mass loss of

$$\Delta M = 6.7 \cdot 10^{26}\,\mathrm{kg} = 3.4 \cdot 10^{-4} M_\odot,$$

▶ During its whole life of 5 billion years up to now the Sun has lost only 0.03% of its mass due to the emitted radiation.

Although accurate measurements have shown that the luminosity of the Sun is not strictly constant in time (Fig. 4.91) but shows small periodic and random variations, many independent indicators confirm that the mean radiation power of the Sun had not changed during the last 5 billion years.

During this period, the total emitted radiation energy amounts to.

$W = 6 \cdot 10^{43}$ J.

The question is now:

What is the source of this enormous energy??

We will check the different possible sources:

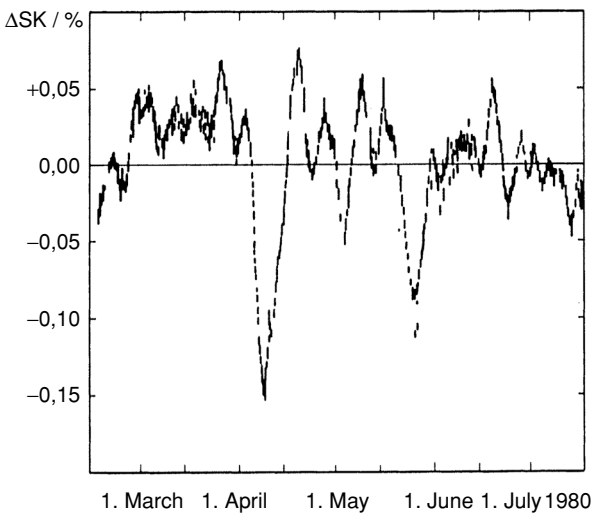


Fig. 4.91 Temporal variation of the solar constant (from J. K Beatty: Die Sonne und ihre Planeten. Physik Verlag Weinheim 1985)

According to the Stefan-Boltzmann radiation law (see Vol. 3 Sect. 3.1.4), the total radiation power of a black body with the surface area $A = 4\pi R_\odot^2$ at the temperature T is

$$\frac{dW}{dt} = L_\odot = 4\pi R_\odot^2 \cdot \sigma \cdot T^4, \tag{4.42}$$

where $\sigma = 5.670367 \cdot 10^{-8}\,\mathrm{W} \cdot m^{-2} \cdot T^{-4}$ is the Stefan-Boltzmann constant.

If we regard the Sun as a black body (which is justified as a good approximation) we obtain the effective surface temperature

$$T_{\mathrm{eff}} = \left(\frac{L_\odot}{4\pi\sigma R_\odot^2}\right)^{1/4}. \tag{4.43}$$

The effective temperature T_{eff} of the Sun is the temperature of a spherical black body with Radius $R_\odot$, which emits the same radiation power as the Sun.

Inserting into (4.43) the numerical values of $L_\odot, R_\odot$ and σ one obtains the effective surface temperature of the Sun

$$\boldsymbol{T_{\mathrm{eff}}} = 5770\,\mathrm{K}.$$

The temperature rises towards the inner part very strongly. Sun models which will be discussed later on assume a mean temperature $\overline{T}$ of about 5 million Kelvin. The thermal energy content W of a sphere with the mean temperature $\overline{T}$*is*

$$W = N \cdot k \cdot \overline{T}$$

where N = M/m is the number of particles with mass m in the Sun with mass M. Inserting the numerical values gives

$$W \approx 2 \cdot 10^{41}\,\mathrm{Ws}.$$

This is only 0.3% of the energy radiated over 5 billion years. Since both the Sun radius and the luminosity $L_\odot$ have not changed over the last 3–5 billion years, the radiation energy cannot have been supplied by the thermal energy of the Sun but must need an energy source in the inner part of the sun which delivers sufficient energy to maintain the luminosity of the sun. This energy must be continuously transported from the inner part to the surface.

Another energy source could be the rotational energy of the Sun, which has a rotation period of $T_{\mathrm{rot}} = 25.38\,\mathrm{d} \Rightarrow$.

$$\omega = \frac{2\pi}{T_{\mathrm{rot}}} = 2.8 \cdot 10^{-6}\,\mathrm{s}^{-1} \Rightarrow E_{\mathrm{rot}} = \frac{1}{2} M \cdot \omega^2 = 4 \cdot 10^{16}\,\mathrm{Ws}.$$

This shows that the rotational energy is completely negligible, compared with the thermal energy.

4.6.2 Mean Values of Temperature and Pressure Inside the Sun

We will at first try to estimate temperature and pressure inside the Sun using a simplified model. If we regard the Sun as homogeneous sphere the potential energy due to the gravitational attraction is (see Problem 4.10)

$$E_{\mathrm{pot}} = -\frac{3}{5} G \cdot \frac{M_\odot^2}{R_\odot} = -2.35 \cdot 10^{41}\,\mathrm{J}. \tag{4.44}$$

The mean kinetic energy can be obtained using the **Virial theorem**, which states that for interaction forces $F \propto r^{-n}$ between the interacting particles the relation between mean potential and mean kinetic energy is

$$\overline{E_{\mathrm{kin}}} = -\left(\frac{1}{n}\right) \cdot \overline{E_{\mathrm{pot}}}. \tag{4.45}$$

Since at the high temperatures inside the Sun all atoms are ionized, the matter consists only of atomic nuclei (protons and helium nuclei) and electrons. The dominant force between the particles is then the Coulomb force with n = 2. Anticipating hydrostatic and thermal equilibrium the Virial theorem states:

$$\overline{E_{\mathrm{kin}}} = \left(\frac{3}{2}\right) \cdot \left(\frac{M}{\overline{m}}\right) \cdot k\overline{T} = -\left(\frac{1}{2}\right)\overline{E_{\mathrm{pot}}} = 1.2 \cdot 10^{41}\,\mathrm{J}, \tag{4.46}$$

where $\overline{m}$ is the mean mass of the particles (nuclei and electrons) and $\frac{M}{\overline{m}} = N$ the number of particles in the Sun. The kinetic energy is the sum of thermal energy $\left(\frac{3}{2}\right)N \cdot kT$ and rotation energy where the latter can be completely neglected as was shown above.

The main part of the Sun material is hydrogen (75%). Inserting into (4.46) for $\overline{m}$ the mass of the hydrogen atom ($m = 1.67 \cdot 10^{-27}$ kg), we obtain for the mean temperature of the Sun the value

$$\overline{T} = \frac{\overline{m}}{3M_\odot \cdot k} E_{\mathrm{pot}} \approx 5 \cdot 10^6 \mathrm{K}. \tag{4.47}$$

At this high temperature, hydrogen is completely ionized. This means that the inner part of the Sun is a plasma of protons and electrons.

Therefore, the mean mass of the particles is

$$\overline{m} = \frac{1}{2}\left(m_{\mathrm{p}} + m_{\mathrm{e}}\right) \approx \frac{1}{2} m_{\mathrm{p}} \tag{4.48}$$

Inserting this into (4.48) gives the mean temperature

$$\boldsymbol{\overline{T}_\odot} = 2.5 \cdot 10^6\,\mathrm{K} \tag{4.49}$$

which differs from (4.47) by the factor 2 because we have now used the value of the mean mass $\overline{m} = \frac{1}{2}m_p$ instead of m_p. Since the temperature at the surface $r = R_\odot$ is only 5700 K, there must be a steep temperature gradient from the surface to the central part of the Sun, and the temperature in the central part at r = 0 must be much higher than the mean temperature. The temperature gradient leads to an energy transport form the inner part to the surface.

For the estimation of the pressure at the central part, a simple coarse estimation is helpful. Imagine the Sun is cut into two half spheres, which attract each other due to the gravitational attraction with the force

$$F_G = -G \cdot \frac{(M_\odot/2)^2}{r^2} \tag{4.50}$$

where $r_c = \left(\frac{3}{4}\right)R_\odot$ is the distance between the centers of mass in the two halfspheres. The pressure at the cutting surface πR^2 is then

$$\overline{p} = \frac{F_G}{\pi R_\odot^2} = G \cdot \frac{4M_\odot^2}{9\pi R_\odot^4}. \tag{4.51}$$

Inserting the numerical values gives

$$\overline{p} = 1.6 \cdot 10^{14}\mathbf{Pa}.$$

From the general gas equation

$$\overline{p} \cdot V = N \cdot k \cdot \overline{T} \tag{4.52}$$

we get with $V = M_\odot/\overline{\varrho}_\odot$ and $N = M_\odot/\overline{m}$ the mean temperature $\overline{T} = 7 \cdot 10^6$ K if we assume for the mean mass density of the Sun $\overline{\varrho}_\odot = 1.4\,\mathrm{kg/dm^3}$ and $\overline{m} = \frac{1}{2}m_p$. This estimation, therefore, gives the same order of magnitude for the temperature as the Virial theorem.

Exactly speaking, the general gas equation is only valid for the ideal gas, which demands that the mean kinetic energy is large compared to the mean attractive potential energy between protons and electrons. This condition is indeed fulfilled in the inner part of the Sun, as can be seen as follows.

The potential energy due to the Coulomb force between protons and electrons is

$$E_{pot} = -\frac{e^2}{4\pi\varepsilon_0 \cdot \overline{d}} \tag{4.53}$$

where $\overline{d}$ is the mean distance between the particles. This mean distance can be obtained from

$$\overline{d} = \sqrt[3]{V_\odot/N} = \sqrt[3]{\frac{\overline{m}}{\varrho_\odot}} \approx 10^{-10}\,\mathrm{m}. \tag{4.54}$$

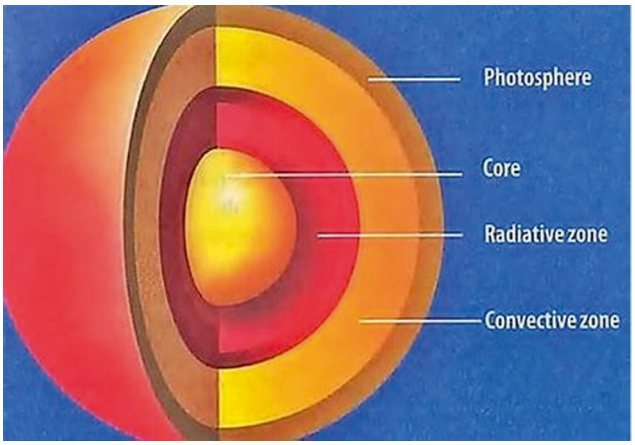


Fig. 4.92 Radial structure of the Sun (Wikipedia)

The condition for the ideal gas is then

$$\overline{E}_{kin} = 2 \cdot \frac{3}{2}k\overline{T} \gg \frac{e^2}{4\pi\varepsilon_0\overline{d}}. \tag{4.55}$$

Inserting $\overline{T} = 3 \cdot 10^6$ K and $\overline{d} = 10^{-10}$ m one obtains

$$\overline{E}_{kin}/\overline{E}_{pot} \approx \frac{1.2 \cdot 10^{-16}\mathrm{J}}{2.3 \cdot 10^{-18}\,\mathrm{J}} = 52 \tag{4.56}$$

which proves that the particles in the inner part of the Sun can be treated as ideal gas.

In order to determine energy production and energy transport more quantitatively, we have to improve these crude models and we should try to calculate the radial dependence of pressure and temperature in the inner part of the Sun which needs a more realistic model of the structure of the Sun (Fig. 4.92).

4.6.3 Radial Dependence of Pressure, Density and Temperature

In order to develop a more realistic model of the Sun, we regard it as hot spherical symmetric gas ball. Due to gravitation, the outer layers exert pressure onto the inner layers, which implies that the pressure p(r) increases with decreasing distance r from the center of the Sun. This is completely analog to the situation in the Earth atmosphere (see Vol. 1. Sect. 7.6).

The mass element in Fig. 4.93

$$\mathrm{d}m = \varrho(r) \cdot \mathrm{d}V = \varrho(r) \cdot \mathrm{A(r)dr}$$

at the distance r from the center experiences the gravitational force

$$\mathrm{d}F_G(r) = G \cdot \frac{M(r)}{r^2}\mathrm{d}m = G \cdot \frac{M(r)}{r^2}\varrho(r)\mathrm{d}A\mathrm{d}r \tag{4.57}$$

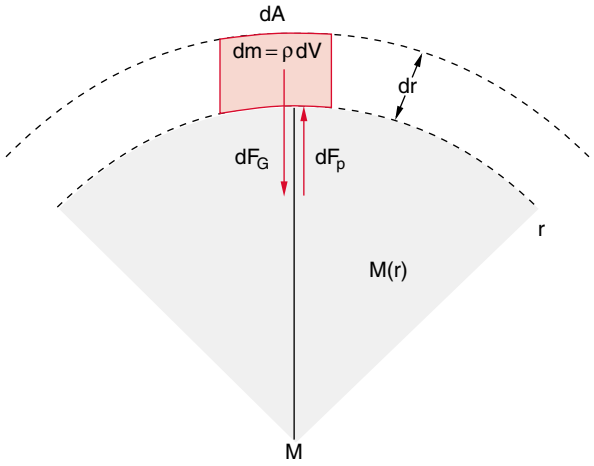


Fig. 4.93 Determination of the radial pressure progression p(r) in the Sun

where M(r) is the mass inside the sphere with radius r (Fig. 4.93). The gas layer between the radii r and $r + \mathrm{d}r$ is only stable if this gravitational force is compensated by the pressure force

$$\begin{aligned} \mathrm{d}F_p &= \big[p(r+\mathrm{d}r) - p(r)\big]\mathrm{d}A \\ &= -\frac{\mathrm{d}p}{\mathrm{d}r}\mathrm{d}r\mathrm{d}A \end{aligned} \tag{4.58}$$

This gives the relation

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -G \cdot \frac{M(r)}{r^2}\varrho(r) \tag{4.59}$$

between pressure gradient $\mathrm{d}p/\mathrm{d}r$ and density $\varrho(r)$. The mass of the spherical shell between the radii r and s $r + \mathrm{d}r$ is

$$\mathrm{d}M = 4\pi r^2 \varrho(r)\mathrm{d}r$$

which gives

$$\Rightarrow \frac{\mathrm{d}M}{\mathrm{d}r} = 4\pi r^2 \varrho(r). \tag{4.60}$$

We can also derive the relation for the energy equilibrium condition.

For a stationary equilibrium, the total energy produced in the inner part of the Sun has to be transported towards the surface, from where it is emitted as radiation. The energy flux through a spherical surface A® i®(r). If inside a spherical shell with thickness dr, the power

$$\frac{\mathrm{d}P}{\mathrm{d}r} = \varepsilon \cdot \mathrm{d}M(r) = 4\pi r^2 \varepsilon \cdot \varrho(r)\mathrm{d}r \tag{4.61}$$

is produced. The energy conservation demands with (4.60), the relation

$$\mathrm{d}L_r = L(r+\mathrm{d}r) - L(r) = \varepsilon \cdot \mathrm{d}M(r) \tag{4.62}$$

and we obtain

$$\frac{\mathrm{d}L_r}{\mathrm{d}r} = 4\pi r^2 \varepsilon \cdot \varrho(r) \tag{4.63}$$

where $\varepsilon(\varrho, T)$ is the power produced per unit mass at the distance r from the center of the Sun which depends on the density ϱ as well as on the temperature T.

The radial temperature progression depends on the specific energy transport, which can be realized by radiation, heat conduction or convection. We will see in Sect. 4.6.6 that in the inner part of the Sun ($0 \le r \le 0.8 R_\odot$), the energy transport occurs nearly exclusively by radiation. Then the temperature gradient becomes (see Problem 4.11)

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -C \cdot \frac{L(r)}{4\pi r^2}\frac{\kappa}{T^3} \cdot \varrho(r) \tag{4.64a}$$

where κ gives the absorption of the radiation per unit mass (opacity), and C is a constant, which depends on L(r) and $\varrho(r)$.

In case of convective energy transport (for r>0.8 $R_\odot$), one obtains the adiabatic temperature gradient

$$\frac{\mathrm{d}T}{\mathrm{d}r} = \frac{c_p - c_V}{c_p}\frac{T}{p} \cdot \frac{\mathrm{d}p}{\mathrm{d}r} \tag{4.64b}$$

where c_p and c_V are the specific heats (see Vol. 1 Sect. 10.1.5).

Eqs. (4.59)–(4.64) relate the radial gradients of pressure p(r), mass M(r), temperature T(r) and energy flux L(r) with the radial density progression $\varrho(r)$.

We have seen above that the gaseous matter in the Sun can be treated as ideal gas. Therefore, the relation between pressure p(r) and density $\varrho(r)$ is the equation of state

$$p(r) = \frac{kT}{\overline{m}}\varrho(r) \tag{4.65}$$

The solution of the differential Eqs. (4.59)–(4.64) has to consider the following boundary conditions:

- The mass $M(r = R_\odot)$ must be the Sun mass $M_\odot$.
- $M(r = 0) = 0$.
- There is no point-like energy source at the center, i.e. $L(r = 0) = 0$.

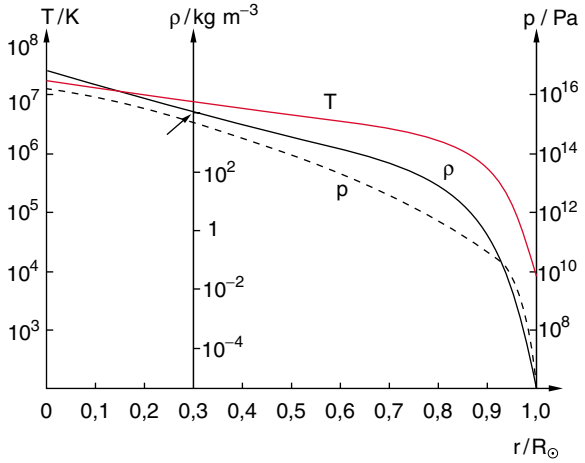


Fig. 4.94 Radial dependence of temperature T(r), pressure p(r) and density ρ(r) in the Sun (after R. Kippenhahn, A Weigert: Stellar Structure and Evolution, Springer Heidelberg 1990)

– The values for temperature $T(R_\odot)$ and pressure $p(R_\odot)$ at the surface of the Sun are known (see Problem 4.12). Since they are very small compared to T(r=0) and p(r=0) we can set in a good approximation $T(R_\odot) = 0$and $p(R_\odot) = 0$.

Since the chemical composition of the Sun material is known, the functions p(r), T(r) and L(r) with their boundary conditions can be calculated [20]. They are plotted in Fig. 4.94. These functions yield for $r = R_\odot$ the total mass $M(R_\odot)$ of the Sun, and the luminosity $L(R_\odot)$. These calculated values can be compared with the measured quantities and can therefore prove the quality of the calculations.

For r=0, the central temperature becomes

$$T(r=0) = 15 \cdot 10^6\,\mathrm{K}.$$

This value can be also obtained by a simple estimation: Assuming a linear temperature progression

$$T(r) = T_0(1 - r/R_\odot) \tag{4.66}$$

The mean temperature is

$$\overline{T} = \frac{1}{V_\odot}\int_0^{V_\odot} T \cdot \mathrm{dV}$$

$$= \frac{3T_0}{4\pi R_\odot^3}\int_0^{R_\odot}(1 - r/R_\odot)4\pi r^2 \mathrm{d}r = \frac{1}{4}T_0. \tag{4.67}$$

Since we had already estimated a value $\overline{T} = (3{-}4) \cdot 10^6 K$ from the general gas equation, we get the central temperature $T(r=0) = 4 \cdot \overline{T} = (12{-}16) \cdot 10^6$ K, which shows that the crude estimation was not too bad.

4.6.4 Energy Production Inside the Sun

Since at a temperature of $T = 10^7$ K, all atoms are completely ionized, chemical processes can be excluded as energy source because they are based on the formation of molecules from atoms.

The gravitational energy, which is released when a giant gas cloud with radius $R \gg R_\odot$ contracts to form the Sun, is

$$E_0 = E_{\mathrm{pot}}(R) - E_{\mathrm{pot}}(R_\odot) = GM_\odot^2\left(\frac{1}{R} - \frac{1}{R_\odot}\right)$$
$$\approx +G \cdot M_\odot^2/R_\odot, \tag{4.68}$$

because $R \gg R_\odot$. Inserting the numerical values gives

$$E_0 = 3.8 \cdot 10^{41}\,\mathrm{J}.$$

As has been estimated in Sect. (4.6.1), the Sun has emitted during its present lifetime the total radiation energy of $6 \cdot 10^{43}$ J, which is 180 times more than the gravitational energy, which therefore cannot supply this emitted radiation energy.

▶ After having excluded the thermal energy and the rotational energy of the Sun, the only possible energy source which must be regarded are nuclear fusion processes. Since the Sun consists mainly of hydrogen gas (H=73%, He=25%, other elements < 2%), the fusion of hydrogen to Helium must play the major role for the energy production in the Sun.

For these processes, the following possibilities exist:

4.6.4.1 The p-p Chain

$$^1_1\mathrm{H} + ^1_1\mathrm{H} \rightarrow ^2_1\mathrm{D} + \mathrm{e}^+ + \nu_\mathrm{e} + \Delta E(1.19\,\mathrm{MeV}) \tag{4.69a}$$

The deuteron D can be also produced by the reaction (Fig. 4.92)

$$\mathrm{p} + \mathrm{p} + \mathrm{e}^- \rightarrow ^2_1\mathrm{D} + \nu_\mathrm{e} + \Delta E(1.44\,\mathrm{MeV}) \tag{4.69b}$$

In the following step, the deuteron can fuse to the helium isotope $^3_2$He

$$^2_1\mathrm{D} + ^1_1\mathrm{H} \rightarrow ^3_2\mathrm{He} + \gamma + \Delta E(5.49\,\mathrm{MeV}) \tag{4.69c}$$

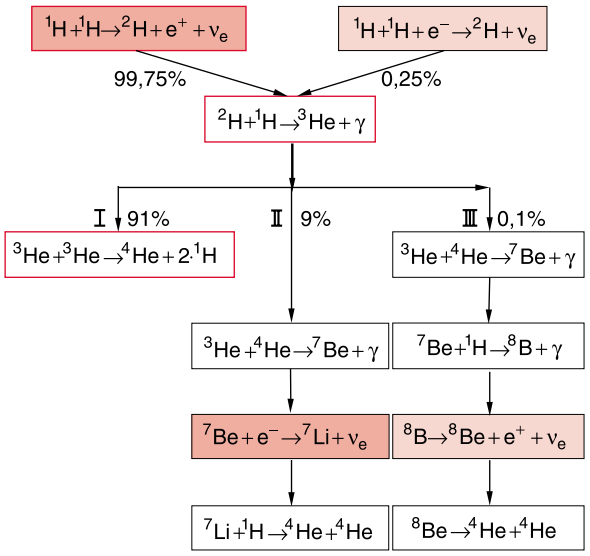


Fig. 4.95 The p-p fusion chain. All reaction steps where Neutrinos are generated are red underlined

The fusion to $^4_2$He can now proceed in several ways (Fig. 4.95):

The major part follows the reaction

$$\text{I}: \ {}^3_2\text{He} + {}^3_2\text{He} \rightarrow {}^4_2\text{He} + \text{p} + \text{p} + \Delta E(12.85\,\text{MeV}) \tag{4.69d}$$

In this reaction (chain (4.69a–d)) altogether four protons are converted into one $^4_2$He nucleus. The total energy released during these processes is

$$\begin{aligned}\Delta E &= (2 \times 1.19 + 2 \times 5.49 + 12.85)\text{MeV} \\ &= 26.2\,\text{MeV},\end{aligned}$$

which corresponds to 6.55 MeV per proton.

A smaller part of the fusion reaction proceeds via the production of beryllium

$$^3_2\text{He} + ^4_2\text{He} \rightarrow {}^7_4\text{Be} + \gamma \tag{4.70a}$$

where two reaction paths occur:

$$\text{II}: \left\{ \begin{matrix} {}^7_4Be + \text{e}^- \rightarrow {}^7_3Li + \nu_\text{e} \\ {}^7_3Li + p \rightarrow {}^4_2He + {}^4_2He \end{matrix} \right\} + \Delta E(25.67\,\text{MeV}), \tag{4.70b}$$

$$\text{III}: \left\{ \begin{matrix} {}^7_4Be + p \rightarrow {}^8_5B + \gamma \\ {}^8_5B + \text{e}^- \\ {}^8_4Be + \nu_\text{e} \rightarrow {}^4_2He + {}^4_2He \end{matrix} \right\} + \Delta E(19.28\,\text{MeV}). \tag{4.70c}$$

The reaction rates $R/\left(\text{m}^{-3}\text{s}^{-1}\right)$ for the fusion processes

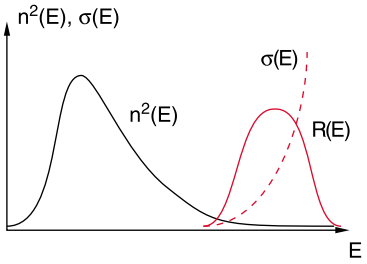


Fig. 4.96 Schematic progression of cross-section σ(E), energy distribution n^2(E) and fusion rate R(E) as a function of the thermal energy E (i.e. the temperature T)

$$R = n_1 \cdot n_2 \cdot \bar{v}_\text{r} \cdot \sigma(v_\text{r}) \tag{4.71}$$

are proportional to the product $n_1 \cdot n_2$ of the densities of the collision partners, the mean relative velocities $\bar{v}_r$ of the collision partners and the collision cross section $\sigma(v_\text{r})$, which also depends on the relative velocity v_r. In Fig. 4.96 the energy dependences of $n^2(E), \sigma(E)$ and R(E) are plotted. The figure illustrates that the fusion processes are mainly induced by collision partners with the highest kinetic energy from the high energy tail of the Maxwellian distribution, where the density n of the particles is already very small. We should be happy about this fact because it restricts the fusion processes to a small fraction of the Sun material and keeps the fusion rate at a moderate level. Otherwise, we would suffer from a gigantic nuclear explosion of the Sun. Under the given facts, the nuclear fusion rate stabilizes at a certain equilibrium level. An increase of the fusion rate would increase the temperature but decrease the density. The net effect is a decrease of the fusion rate.

Since the energy production must supply the radiation losses, this implies that the fusion power must be $4 \cdot 10^{26}$ W. This corresponds to a fusion rate of $4 \cdot 10^{38}$ protons per second.

Dividing this by the total number of protons inside the core of the Sun ($r \leq 0.5\ R_\odot \rightarrow V_\text{core} = (1/8)\ V_\odot$) where the fusion processes occur, the power density is only 0.4 mW/kg, which is less than the power density in our body necessary to maintain the body temperature. **An astonishing result**!!

4.6.4.2 Production of Carbon Nuclei

If a sufficient amount of helium has been produced by the p-p chain, the production of carbon nuclei can proceed by a triple process: Three helium nuclei (α-particles) produce a carbon nucleus (Fig. 4.97).

At first two α-particles collide with each other and form a Beryllium nucleus according to the reaction

$$^4_2\text{He} + ^4_2\text{He} \rightarrow {}^8_4\text{Be} + \text{y}(\Delta E = -0.09\,\text{MeV}) \tag{4.72a}$$

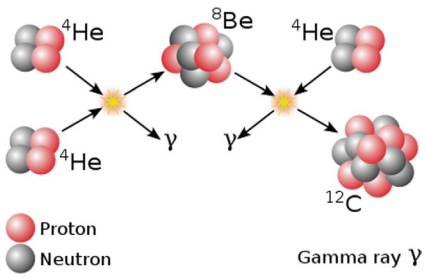


Fig. 4.97 The triple α-process (Wikipedia)

The Berylium nucleus is highly unstable and decays within $8 \cdot 10^{-17}$ s into smaller nuclei. However, if it collides within this short time with a helium nucleus $^4_2$He, the stable carbon nucleus $^{12}_6$C can be produced:

$$^4_2\mathrm{He} + ^8_4\mathrm{Be} \rightarrow ^{12}_6\mathrm{C} + \mathrm{y}(\Delta E = +7.37\mathrm{MeV}) \tag{4.72b}$$

4.6.4.3 The CNO Cycle

There is another fusion process where protons are involved, which, however, only plays a minor role in our Sun but becomes the main process in stars with higher core temperatures. Here, the fusion of protons to $^4_2$He nuclei is catalyzed by carbon nuclei according to the CNO-fusion cycle (postulated by Hans Bethe and Carl Friedrich von Weizsäcker) (Fig. 4.98). The following reaction occurs:

$$^{12}\mathrm{C} + ^1\mathrm{H} \rightarrow ^{13}\mathrm{N} + \gamma$$

$$^{13}\mathrm{C} + \mathrm{e}^+ + \nu_\mathrm{e}, \leq \tag{4.73a}$$

$$^{13}\mathrm{C} + ^1\mathrm{H} \rightarrow ^{14}\mathrm{N} + \gamma, \tag{4.73b}$$

$$^{14}\mathrm{N} + ^1\mathrm{H} \rightarrow ^{15}\mathrm{O} + \gamma$$

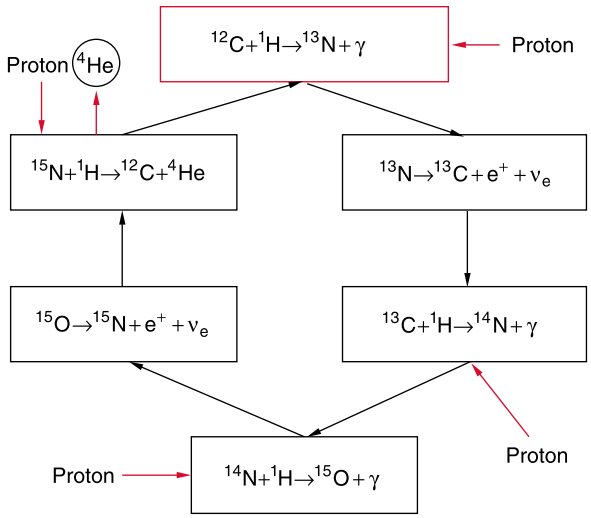


Fig. 4.98 The CNO cycle

$$^{15}\mathrm{C} + \mathrm{e}^+ + \nu_\mathrm{e}, \tag{4.73c}$$

$$^{15}\mathrm{N} + ^1\mathrm{H} \rightarrow ^{12}\mathrm{C} + ^4\mathrm{He}. \tag{4.73d}$$

The carbon nuclei act only as catalysator for the fusion of four protons to form a $^4_2$He nucleus. This process plays a role only if sufficient carbon is present in the fusion zone.

The slowest reaction is (4.73c). It limits the reaction rate of the whole CNO process. The temperature dependence of the two reactions p-p fusion and CNO cycle is plotted in Fig. 4.99a while Fig. 4.99b shows the dependence of the two processes on the distance r from the center. One realizes that the CNO cycle dominates for temperatures above 18 Million Kelvin and is therefore restricted to the inner core of a star.

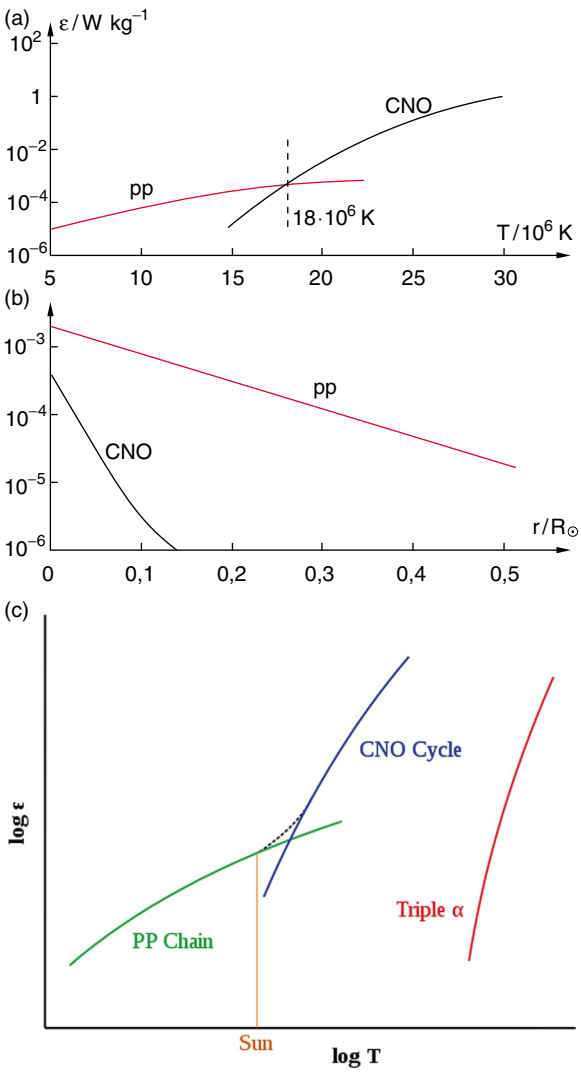


Fig. 4.99 **a** Temperature dependence of the specific energy production ε for the p-p fusion chain and the CNO-cycle. **b** Specific energy production in the Sun for the two fusion processes as a function of the distance r/R from the Sun center. **c** Comparison of the temperature dependence of the three fusion processes (Wikipedia)

In both processes, four protons form one $^4_2$He nucleus according to the reaction balance

$$4 \cdot {}^1_1\mathrm{H} \rightarrow 1 \cdot {}^4_2\mathrm{He} + 2\mathrm{e}^+ + 2\nu_\mathrm{e} + 2\gamma + \Delta E \qquad (4.74)$$

Per fusion cycle, the energy $\Delta E = 26.21\,\mathrm{MeV} = 4.1 \cdot 10^{-12}$ J is released for the p-p cycle, whereas in the CNO cycle, the corresponding energy is $\Delta\mathrm{E} = 25.03$ meV.

In order to supply the emitted radiation energy solely by the p-p cycle, the number of fusion process per sec must be

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{L_\odot}{\Delta E} = \frac{3.82 \cdot 10^{26}\,\mathrm{W}}{4.1 \cdot 10^{-12}\,\mathrm{J}} \approx 10^{38}\,\mathrm{s}^{-1} \qquad (4.75)$$

▶ If the Sun would use only 5% of its hydrogen content, the fusion energy of the p-p cycle would be sufficient to supply the emitted radiation power for 10^{10} years. Since its present lifetime is 5 billion years, it can therefore keep its present status at least for further 5 billion years.

In Fig. 4.99c, the three fusion processes and their energy dependence are compared, which illustrates that the triple α-process only plays a significant role at very high temperatures.

4.6.5 The Problem with the Sun-Neutrinos

During the fusion reactions, neutrinos are released, which can leave the Sun unimpeded because of their extremely small interaction with matter. Measurements of their number and energy distribution give direct information about the fusion processes 8 min ago. The time delay $\Delta\mathrm{T} = 500\,\mathrm{s} = 8.3$ min is the travel time of the neutrinos from the Sun to the Earth.

Presently great efforts are undertaken to measure these Sun neutrinos [26, 27]. The flux density of the neutrinos that reach the Earth should be $N(\nu) = 10^{15}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$, if the generally accepted Sun model is correct.

In Fig. 4.100, the theoretical neutrino spectrum is shown, which is expected from the fusion processes (4.69). There are continuous spectra and discrete line spectra. The continuous spectra are plotted in the units $[\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{MeV}^{-1}]$ (left scale), the line spectra in the unit $[\mathrm{m}^{-2}\,\mathrm{s}^{-1}]$ (right scale).

The neutrinos resulting from the p-p reaction (4.69a) show a continuous energy spectrum with a maximum energy of 0.4 MeV because the positron e^+ and the neutrino ν_e share the reaction energy (see Vol. 4 Sect. 3.4 ß-decay). The neutrinos from the p-p-reaction (4.69b) are mono-energetic with the energy $E = 1.44$ MeV, as well as the neutrinos from the two-body decay of the $^7_4$Be (4.70b), which appear in two groups with energies E_1 and E_2 depending on

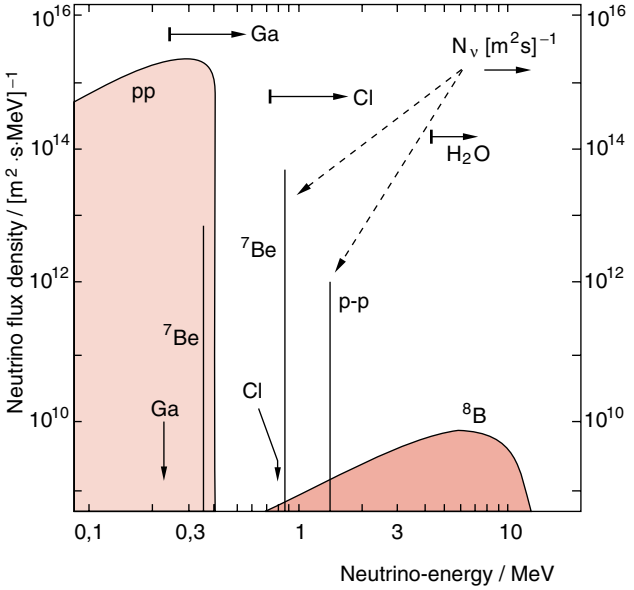


Fig. 4.100 Theoretical Sun-neutrino spectrum as a function of the neutrino energy

Fig. 4.101 Raymond Davis

the excitation level of the ^{7}Li-nucleus. The highest energy has the neutrinos from the three-body decay of the ^{8}B nucleus (process III 4.71c). However, this process contributes only 0.1% of all fusion processes in the Sun.

Extensive measurements of the rate of Sun neutrinos have been performed by R. Davis (Nobel Prize 2002) (Fig. 4.101) in an abandoned goldmine in Dakota [24]. The neutrinos passed through a large tank with 610 tons of Tetra-chlorine ethylene $\mathrm{C_2Cl_4}$ (Fig. 4.102) where they produced radioactive Argon according to the reaction

$$\nu_\mathrm{e} + {}^{37}\mathrm{Cl} \rightarrow {}^{37}\mathrm{Ar} + \mathrm{e}^- \qquad (4.76)$$

$$^{37}\mathrm{Ar} \xrightarrow[\text{electron capture}]{35\mathrm{d}} {}^{37}\mathrm{Cl} + \text{Auger} - \text{electrons}$$

The gaseous argon has been extracted from the liquid chlorethylene and its decay counted by special counting devices.

Fig. 4.102 Tank filled with C_2Cl_4 in the Homestak Gold-mine in South-Dakota (Wikipedia)

The result of these careful measurements was a neutrino rate of 0.5 Neutrinos/day. This is only 1/3 of the rate expected from realistic Sun models. However, with this technique only neutrinos with energies E>0.81 MeV could be detected. Nevertheless, the Sun experts were very much disturbed by this result. Therefore, several labs tried to improve the neutrino measurements.

One of these experiments is the GALLEX experiment in a side adit of the motorway tunnel in the Grand Sasso Mountains in Italy. Here, a large Gallium tank (30 tons of ($GaCl_3$) is used for the capture of neutrinos. The reaction

$$^{71}\mathrm{Ga} + \nu \rightarrow {}^{71}\mathrm{Ge} + \mathrm{e}^- \qquad (4.77)$$

$$^{71}\mathrm{Ge} \xrightarrow[\text{electron capture}]{11.4\mathrm{d}} {}^{71}\mathrm{Ga} + \text{Auger - electrons}$$

has a threshold energy of only 0.23 MeV and can therefore also detect the neutrinos of the p-p fusion reaction (4.69a), which is much more frequent (Fig. 4.97).

The Germanium atoms in the Gallium liquid are extracted by a special chemical separation process and transferred into the gaseous GeH_4 molecules, which are fed into an Auger-electron-counting device.

In both experiments, the Auger electrons emitted at the decay of ^{37}Cl resp. ^{71}Ge are detected.

The results obtained up to now (Fig. 4.104) clearly demonstrate that indeed a distinct neutrino deficiency was found, which, however, no longer amounts to a factor of 1/3 but rather ½. This is still far outside the error limits. The question is now: What is the reason for this deficiency??

There are two possible causes: Either the Sun model is not correct, which would affect the fusion rates, or the neutrinos change their identity on their way from the Sun interior to the Earth. We will see, that the second cause is correct.

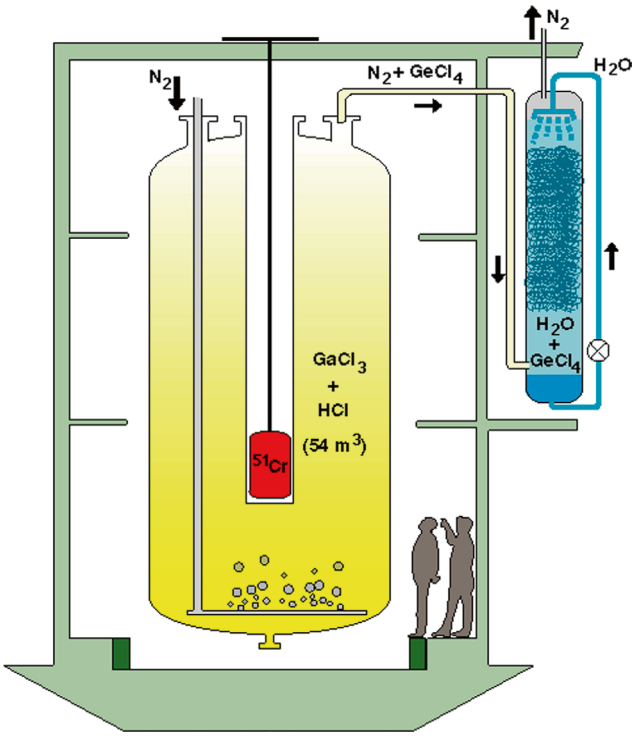


Fig. 4.103 The GALLEX detection tank filled with $GaCl_3$ (https://lappweb.in2p3.fr/neutrinos/nexp.htm)

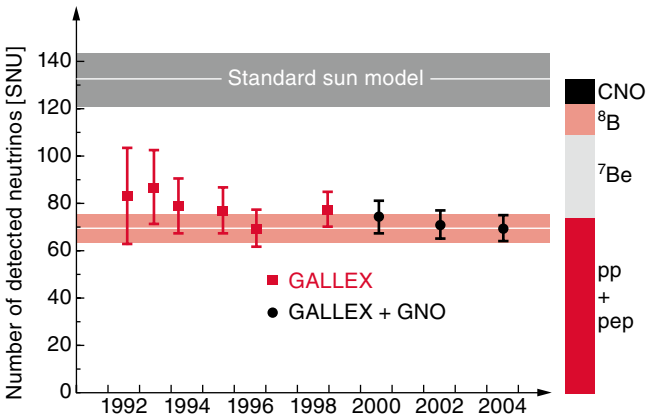


Fig. 4.104 Measured values from the GALLEX experiment for the particle flux of the Sun neutrinos compared with the standard model of the Sun [28]

The neutrinos emitted in the fusion reactions (4.69) are all electron neutrinos ν_e. They change on their way to the Earth into μ-neutrinos, which could not be detected neither in the homestake nor in the GALLEX experiment. A new experimental setup, the Super-Kamiokande (Fig. 4.105), can detect both kinds of neutrinos. It consists of a huge tank with 50.000 tons of water, where the walls are covered by 11.146 photo-multipliers. They detect the Cherenkov radiation, which is emitted by charged particles, which pass

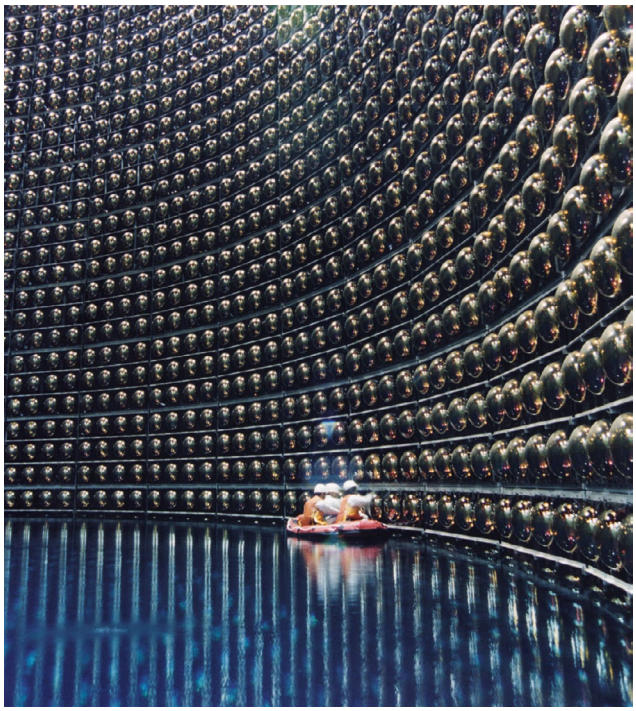


Fig. 4.105 Inner part of the Super-Kamiokande. Inspection of the photomultipliers. During operation, the whole volume is filled with water

through the water with velocities $v > c = c_0/n \approx 0.7c_0$ higher than the velocity of light in water. The water tank is located in a lead-Zinc mine in 1000 m below ground in Japan. The spatial arrangement of the detectors allows the determination of the incident direction of the charged particles. The density of the particle tracks gives information about the type of particles and their energy.

The charged particles are generated by the neutrino-induced reactions in the water tank

$$\nu_e + {}^2_1\mathrm{D} \rightarrow e^- + p + p \quad \text{(conversion of a neutron into a proton)} \tag{4.78a}$$

$$\nu_e + e^- = \nu_e + e^- \quad \text{(elastic scattering of neutrinos at atomic electrons).} \tag{4.78b}$$

Now the neutrino rate from two opposite directions is measured (Fig. 4.103). In one direction, the neutrinos have to pass through the Earth, in the other direction (12 h later) not. Besides the solar neutrinos, there are also neutrinos that are generated by the cosmic radiation in the Earth atmosphere, according to the reactions

$$\pi^+ \rightarrow \nu_\mu + \mu^+ \tag{4.79a}$$

$$\mu^+ \rightarrow e^+ + \overline{\nu}_\mu + \nu_e \tag{4.79b}$$

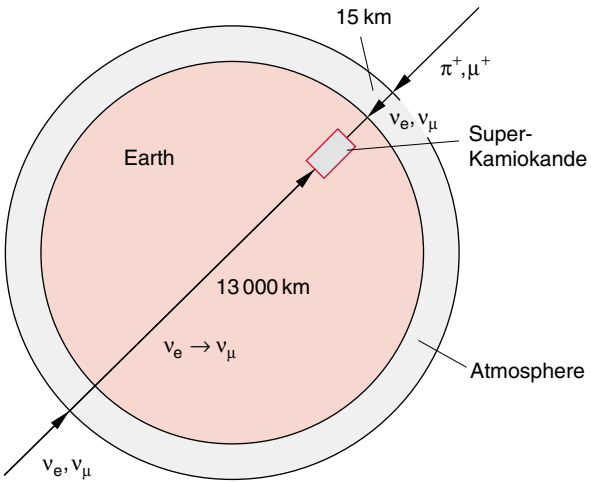


Fig. 4.106 Measuring principle of the Super-Kamiokande detector for the proof of neutrino oscillations

this implies that the neutrino species ν_e, ν_μ *and* $\overline{\nu}_\mu$ are generated. The ratio of the numbers

$$(\nu_\mu + \overline{\nu}_\mu)/\nu_e = 2$$

should be 2. If the ν_e-neutrinos on their way through the Earth (13.000 km) are converted into ν_μ-neutrinos. This ratio should be different for the two detection directions in Fig. 4.106. The probability that a ν_e neutrino converts into a ν_μ-neutrino on its way through the Earth is

$$P(x) = 1 - \sin^2(2\theta) \cdot \sin^2(1.27(\Delta mc^2/E)^2 \cdot x) \tag{4.80}$$

where $\Delta m = m(\nu_\mu) - m(\nu_e)$ is the mass difference between the two neutrino species and E is their energy [28].

During the first measurement period, 4353 events have been detected. The analysis shows that ν_e-neutrinos can convert into ν_μ-neutrinos and back (neutrino oscillations).

▶ **The results of Super-Kamiokande prove that the missing neutrinos from the Sun can be explained by the neutrino oscillations. One does not need a new Sun model.**

4.6.6 The Energy Transport in the Sun

As illustrated in Fig. 4.98, the energy production due to fusion processes occurs essentially inside a sphere with radius $r < 0.3R_\odot$. The energy produced in the inner part of the Sun, but radiated from the Sun surface, has to be transported, i.e. it has to flow through a spherical shell with radius $r = R_\odot/3$. As shown in Fig. 4.94, the temperature gradient through a shell with $\Delta r = R_\odot/10$ at $r = R_\odot/3$ is about

$$\frac{\Delta T}{\Delta r} = 6 \cdot 10^6 K/(R_\odot/10).$$

This corresponds to a radial temperature gradient of

$$\frac{dT}{dr}\left(r = \frac{1}{3}R_\odot\right) = 1.6 \cdot 10^{-2}\,\mathrm{K/m}. \qquad (4.81)$$

With a heat conduction coefficient λ of the Sun material, the heat energy transport through the surface A due to heat conduction becomes

$$\frac{dW}{dt} = \lambda \cdot A \cdot \frac{dT}{dr} \qquad (4.82)$$

inserting realistic numerical values $\left(\lambda = 10^8\,\mathrm{W/(m\,K)}\right)$ for the heat conduction of the hot plasma at $r = R_\odot/3$ one obtains only 1% of the heat flux necessary to supply the radiation power, emitted from the Sun surface.

▶ Heat conduction as energy transport from the inner part of the Sun to its surface is negligibly small!!

In the Earth atmosphere, convection plays an essential role for the heat transport. We will therefore consider, whether it also is important in the Sun.

The radial energy transport by convection is realized by rising packets of hot plasma from a zone with radius r to a higher layer with radius r+h, because of their lower density (Fig. 4.107). This phenomenon, which is also present in the Earth atmosphere, is called **thermal lift**.

During its updrive, the hot packet cools down. This cooling is adiabatic, if the rise time is sufficiently short to prevent any energy exchange with its surroundings. It rises as long as its density is smaller than that of its surroundings. This means that

$$\left|\frac{dT}{dh}\right|_{\text{adiabatic}} < \left|\frac{dT}{dh}\right|_{\text{surrounding}}. \qquad (4.83)$$

The temperature gradient of the surroundings must be larger than the adiabatic gradient of the rising plasma packet.

$$\frac{dT}{dr} = \left(1 - \frac{1}{\kappa}\right)\frac{T}{p}\frac{dp}{dr} \qquad (4.84)$$

with $\kappa = c_p/c_v$.

It turns out that his condition is only fulfilled in the outermost layers with $r > 0.84R_\odot$ because of two reasons:

– In these outer layers, the temperature gradient is extraordinarily large (see Fig. 4.94), i.e. the right side of Eq. (4.83) is particularly large.

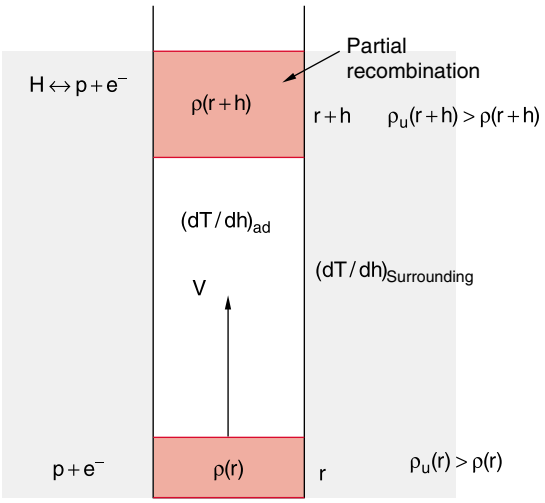


Fig. 4.107 Rising plasma packet with adiabatic cooling and partial recombination

– Part of the plasma particles (electrons and ions) can recombine during the cooling process of the uprising plasma packet. The released recombination energy heats the packet up and defers the cooling process (Fig. 4.107), keeping the left side of Eq. (4.83) small.

▶ The energy transport in the small outer range of $0.84R_\odot < r < 0.98R_\odot$ is mainly realized by convection.

The rise velocity of the plasma packets reaches values up to 1 km/s.

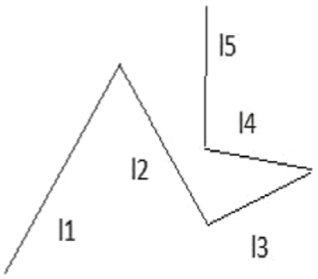
In the predominant part of the Sun ($r < 0.84R_\odot$), **heat radiation** must be responsible for the energy transport.

How is this realized in detail?

The γ-quanta released at the fusion processes (4.69c and 4.73a–c) are either annihilated by pair production (see Vol. 4) or they transfer part of their energy onto the electrons by the Compton effect. The electrons release their energy by *bremsstrahlung*, when they are decelerated during their passage close to protons or He-nuclei. This bremsstrahlung can be again transferred to electrons due to the Compton effect, etc. (Fig. 4.108b). In summary, there is a continuous change between electron energy and photon energy. The photons do not fly straight forward into the radial direction, but suffer at each of these processes a change of their direction (Fig. 4.108a). They diffuse with a very small mean free pathlength through the Sun plasma. The mean free pathlength Λ

$$\Lambda = \frac{1}{n \cdot \sigma} \qquad (4.85)$$

is inversely proportional to the particle density n and the collision cross section σ. It therefore increases with decreasing density, i.e. increasing radius r from the inside of the Sun to its outer regions.



https://lappweb.in2p3.fr/neutrinos/nexp.htm

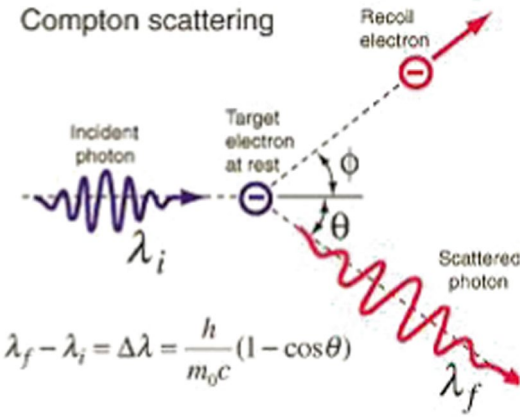


Fig. 4.108 **a** Schematic illustration of photon diffusion in the Sun. **b** Compton scattering. https://lappweb.in2p3.fr/neutrinos/nexp.htm

Example

At $r = 0.1R_\odot$ is $\varrho = 10^5\,\mathrm{kg/m^3} \Rightarrow n = 6 \cdot 10^{31}\,\mathrm{m^{-3}}$ With a scattering cross-section $\sigma = 10^{-26}\,\mathrm{m^2}$ the mean free pathlength becomes $\Lambda = 1.5 \cdot 10^{-6}$ m. In the average, only 1/6 of all scattering processes results in a radiation transport into the outer radial direction. In the range from r=0 to $r = R_\odot/2$ ($3.5 \cdot 10^8$ m), there occur about 10^{15} scattering processes! ◀

It takes in the average 10^7 years for a photon to travel from the central part to the surface, while the direct flight time (without interactions) would be 2 s!

The mean time it takes for a photon to reach the surface of the Sun can be estimated as follows.

The mean thermal energy of the Sun at a mean temperature of $5 \cdot 10^6$ K is

$$W_{\mathrm{thermal}} = \frac{M_\odot}{\overline{m}} \cdot \frac{3}{2} k\overline{T} \approx 1.2 \cdot 10^{41}\,\mathrm{J} \qquad (4.86)$$

The emitted power of the Sun is

$$L_\odot = 3.8 \cdot 10^{26}\,\mathrm{W}$$

The mean energy transport time is then

$$T = W_{\mathrm{thermal}}/L_\odot = 3 \cdot 10^{14}\,\mathrm{s} = 10^7\,\mathrm{years!}$$

This is an astonishingly large value. It means that we get information from the observed Sun radiation about the energy production in the inner part of the Sun, that occurred 10^7 years ago!

Only the neutrinos traverse the Sun essentially unimpeded. It takes them only about 500 s ≈ 8 min to travel from the inside of the Sun to the Earth.

The energy transport in the convection zone takes only a small part of this time. At a thickness $\Delta r = 0.14R_\odot \approx 10^5$ km and a mean velocity of $\overline{v} = 0.5$ km/s of the uprising plasma packets the transport time for the way from $0.84R_\odot$ to $0.98R_\odot$ is about $\overline{t}_{\mathrm{conv}} = \frac{\Delta r}{\overline{v}} = 2 \cdot 10^5\,\mathrm{s} = 55\,\mathrm{h}$.

4.6.7 The Photosphere

When observing the Sun through a strongly attenuating filter or in the evening shortly before sunset one sees a circular disc with sharp edges. This indicates that the visible light comes from a thin layer at the surface of the Sun. The surface brightness L is not constant but decreases slightly with the distance x from the center towards the edges of the Sun disc (Fig. 4.109b). The brightness gradient is steeper for short than for long wavelengths. At the edge, the brightness falls within a small range of 200 km ($2.7 \cdot 10^{-4}R_\odot$) from 40% of its central value to zero.

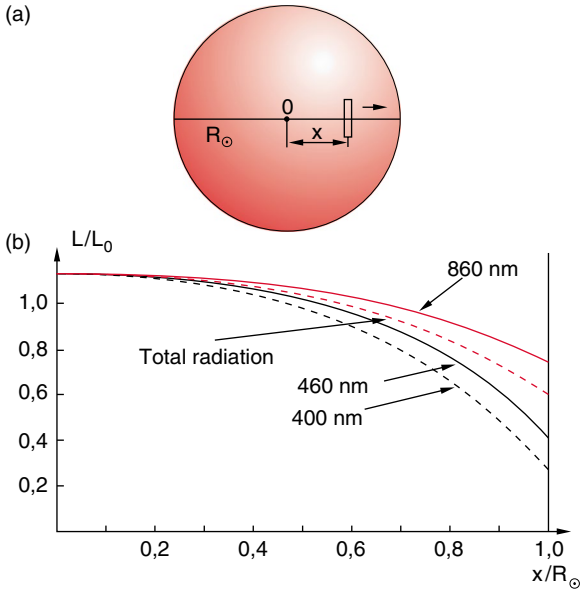


Fig. 4.109 Sun luminosity L(x) as a function of the distance x from the center of the Sun disc. **a** Measuring method. **b** L(x) for different wavelengths λ(nm)

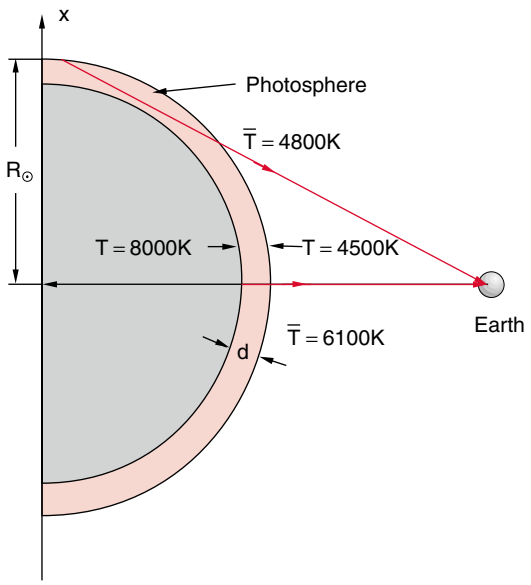


Fig. 4.110 Schematic illustration of the viewing depth into the photosphere for the explanation of the apparent luminosity change L(x). The thickness of the photosphere is here greatly exaggerated

Sliding a small slit in front of the radiation detector across the Sun image in the image plane of a Sun telescope (Fig. 4.109a) the detected radiation power (which is a measure of the brightness) follows the function

$$L(x) = \frac{2}{5}L_0\left(1 + \frac{3}{2}\sqrt{1-(x/R_\odot)^2}\right). \tag{4.87}$$

This gives $L(x=0) = L_0$; $L(x=R_\odot) = \left(\frac{2}{5}\right)L_0$. Defining the mean value $\overline{L}$ of the surface brightness L(x) as

$$\overline{L} = \frac{1}{\pi R_\odot^2}\int_0^{R_\odot} L(x)\cdot 2\pi x\,\mathrm{d}x, \tag{4.88}$$

This gives after inserting (4.87)

$$\overline{L} = \frac{4}{5}L(x=0) \tag{4.89}$$

We can then write (4.87) as

$$L(x) = \frac{1}{2}\overline{L}\left(1 + \frac{3}{2}\sqrt{1-(x/R_\odot)^2}\right). \tag{4.90}$$

These quantitative observations can be explained when assuming that the visible Sun radiation is emitted from a thin surface layer called the **Photosphere**, which has a large absorption coefficient for visible light.

In the central part of the Sun disc, we can look deeper into the photosphere than at the same total absorption in the edges, where the observation direction has a longer path length (Fig. 4.110). Since the temperature of the photosphere decreases from its inner layer to the surface, we see at the center of the Sun disc hotter layers than at the edges. They emit more radiation than the colder layers. This is the reason for the maximum brightness and the higher measured temperature at the center of the Sun disc.

The measured brightness profile L(x) can be converted into a temperature profile T(x), if the Sun is regarded as a black body, for which the Stefan-Boltzmann radiation law $L = \sigma \cdot T^4$ is valid. This assumption is a very good approximation! Taking for the mean surface temperature the value $\overline{T} = T_{\mathrm{eff}} = 5770\,\mathrm{K}$ obtained from measurements of the solar constant (see Sect. 4.6.1) the mean brightness becomes $\overline{L} = \sigma \cdot \overline{T}^4$ and we get with (4.90)

$$T(x) = 5770\,\mathrm{K}\cdot\sqrt[4]{\frac{1}{2} + \frac{3}{4}\sqrt{1-(x/R_\odot)^2}}. \tag{4.91}$$

At the edge of the Sun disc, where we see only into the outer layers of the photosphere (Fig. 4.110) we measure the temperature $T_{\mathrm{outer}} = 5770\cdot(1/2)^4 = 4850\,\mathrm{K}$, whereas in the central part (x = 0) where we receive light from all layers of the photosphere the measured temperature is $T(\mathrm{x}=0) = 6100\,\mathrm{K}$ which represents a mean value from all layers of the photosphere. In Fig. 4.111, the progression of pressure, density and temperature is plotted as a function of the heights h from the inner part of the photosphere, as obtained from more detailed models of the photosphere.

The question is now: How is the continuous spectrum of the Sun generated, although the emission of atoms should

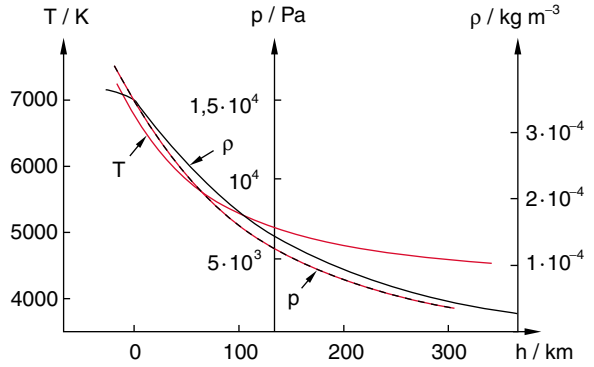


Fig. 4.111 Pressure p(r), density ρ(r) and temperature T(r) in the photosphere

result in a line spectrum and pressure broadening should play no essential role because the pressure in the photosphere is low?

The continuous spectrum cannot be due to transitions between bound states of atoms or ions which would result in discrete line spectra. The situation is different for recombination radiation, where a free electron recombines with an ion or a neutral atom resulting in excited states of a neutral atom or a negative ion (see Vol. 3. Sect. 7.7).

At the mean temperature of the photosphere $\overline{T} = 6000\,\mathrm{K}$ only about 10^{-4} of all H-atoms are ionized. This implies that the recombination of electrons with proton (H^+-ions) is very improbable. However, the recombination with neutral atoms, where negative H^--ions are produced, is much more probable.

The electron affinity of H^- is 0.75 eV. This means that the recombination of an electron with the kinetic energy E_{kin} and a neutral H-atom

$$H + e^- \rightarrow H^- + h \cdot \nu \tag{4.92}$$

generates photons $h \cdot \nu$ with the energy

$$h \cdot \nu = E_{kin}(e^-) + E_B \approx E_{kin}(e^-) + 0.75\,\mathrm{eV} \tag{4.93}$$

The long wavelength limit of this emission is at $h \cdot \nu = 0.75\,\mathrm{eV} \Rightarrow \lambda = 1660\,\mathrm{nm}$. Since E_{kin} follows the temperature dependent Maxwell-Boltzmann-distribution there is no sharp upper limit for $h \cdot \nu$, i.e. no lower limit for the wavelength λ.

The recombination with H^+-ions

$$H^+ + e^- + E_{kin}(e^-) \rightarrow H + h\nu \tag{4.94}$$

plays only a minor role for the visible radiation of the Sun because of the low concentration of H^+-ions. However, it contributes to the ultraviolet part of the Sun spectrum due to the larger ionization energy of the H-atom.

Besides the above-discussed processes of radiative recombination, a radiation continuum can be produced by *Bremsstrahlung* of the electrons in the Coulomb field of the H^+-ions or some less abundant heavy nuclei. In this process at most the kinetic energy of the electrons can be converted into radiation. At the temperature of $T = 6000\,\mathrm{K}$, this gives a short wavelength limit of $\lambda = 1\mu m$. Therefore, this process can only contribute to the infrared emission of the Sun.

Besides the continuous emission, also the inverse process of absorption occurs. The absorption of radiation by negative H^--ions and by excited neutral H-atoms

$$H^- + h\nu \rightarrow H + e^- \tag{4.95a}$$

$$H^* + h\nu \rightarrow H^+ + e^- \tag{4.95b}$$

causes the opacity of the photosphere. The radiation from the inner part of the Sun, entering the photosphere from below, is already absorbed in a small layer of the photosphere.

The H^--ions can loose their electron also by collisions with other species in the photosphere. Both processes lead to the fill up of the reservoir of neutral H-atoms. At longer wavelengths λ, also the process of free-free-absorption

$$e^- + e^- + h \cdot \nu \rightarrow e^- + e^- + \Delta E_{kin}(e^-) \tag{4.96}$$

in the Coulomb field of nuclei as inverse process of the *bremsstrahlung* contributes to the absorption (Fig. 4.112).

The spectral distribution of the Sun radiation is determined by the superposition of absorption and emission. It corresponds to a good approximation to the Planck radiation of a black body with a mean temperature of $\overline{T} = 5770\,\mathrm{K}$.

Observing the continuous emission of the photosphere through a spectrograph with sufficient spectral resolution, one discovers many sharp absorption lines in the bright

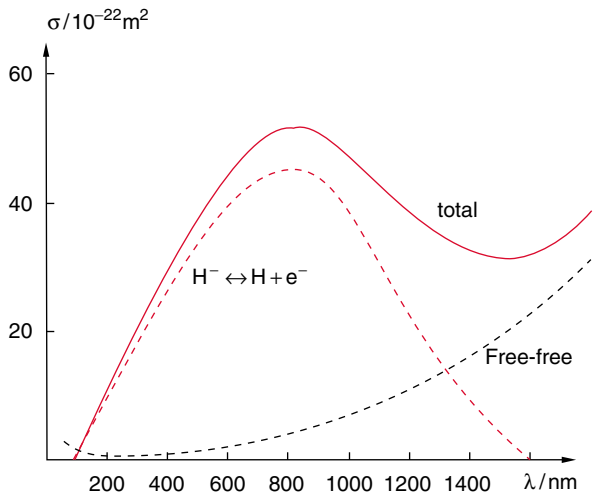


Fig. 4.112 Optical absorption cross-section for photo-detachment $H^- + h \cdot \nu \rightarrow H + e^-(E_{kin})$ and for free-free-transitions of electrons $e^-(E_{kin}) + h \cdot \nu + p \rightarrow p + e^-(E_{kin} + h \cdot \nu)$

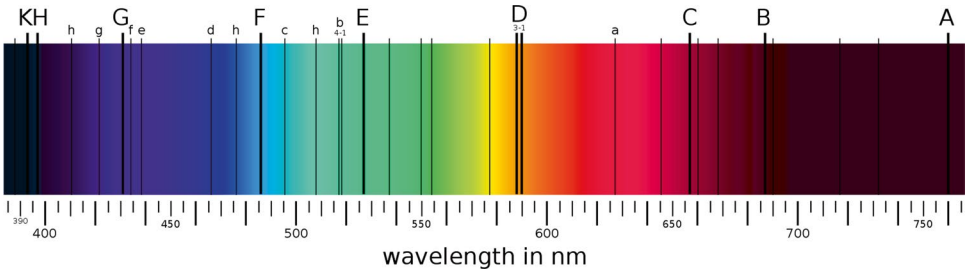


Fig. 4.113 Fraunhofer lines in the Sun spectrum (https://commons.wikimedia.org/w/index.php?curid=7003857)

continuum (Fig. 4.113). They were first observed 1802 by the English physicist William Hyde Wollaston, but are named for the German Optician Joseph Fraunhofer (1787–1826, Fig. 4.114) who had observed the lines in more detail, using prisms for a spectrograph which he had built and polished by himself. He published a compilation of more than 500 lines. For his achievements in optical technology and scientific applications, he was ennobled by the German King.

The Fraunhofer lines result from selective absorption of the continuous radiation from the inner part of the photosphere by atoms in the colder outer part of the photosphere. Since the atoms are located in parts of the photosphere with a lower temperature than in those parts where the continuous radiation is emitted, the absorption exceeds the emission and the lines appear as sharp spectral decreases of the continuous Sun radiation.

The investigation of intensity and linewidth of the Fraunhofer lines gives information about the abundance of the chemical elements in the Sun photosphere and the density of the matter causing the line broadening (Table 4.9).

With the density $n(E_k, h)$ of absorbing atoms in level E_k at the heights h in the photosphere and the optical absorption cross-section σ_{ki} for the transition $k \rightarrow i$, the optical absorption coefficient becomes

$$\alpha(\nu_{ki}) = \left[n(E_k, h) - (g_i/g_k)n(E_i, h)\right] \cdot \sigma_{ki} \qquad (4.97)$$

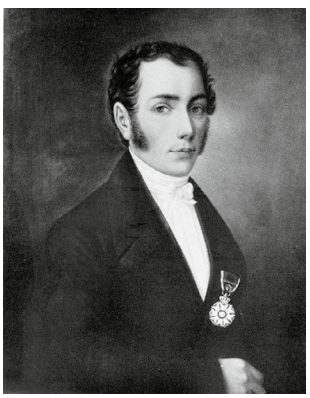


Fig. 4.114 Joseph von Fraunhofer

Table 4.9 Relative abundance of the elements in the Sun atmosphere in atom number densities

H	1.0	Mg	$2.3 \cdot 10^{-5}$
He	0.1	Fe	$3.0 \cdot 10^{-6}$
O	$1.0 \cdot 10^{-3}$	Na	$2.0 \cdot 10^{-6}$
C	$5.2 \cdot 10^{-4}$	Ca	$1.4 \cdot 10^{-6}$
N	$1.0 \cdot 10^{-4}$	Ni	$8.3 \cdot 10^{-7}$
Si	$2.8 \cdot 10^{-5}$	Cr	$2.3 \cdot 10^{-7}$

Radiation with the spectral power density $L(\nu, h_0)$ at the heights h_0 which passes through a layer of absorbing atoms is attenuated by absorption. The transmitted power density at the center of the absorption line is then

$$L_t(\nu_{ki}) = L_0(\nu_{ki}) \cdot e^{-\int_{h_0}^{\infty} \alpha(\nu_{ki}, h)\mathrm{d}h} \qquad (4.98)$$

The absorbed power on the frequency ν_{ki} is then

$$\begin{aligned} \Delta L(\nu_{ki}) &= L_0(\nu_{ki}) - L_t(\nu_{ki}) \\ &= L_0(\nu_{ki})\left(1 - e^{-\int \alpha(\nu_{ki}, h)\mathrm{d}h}\right). \end{aligned} \qquad (4.99)$$

Integration over the line profile of the absorption line gives the missing radiation power of the Fraunhofer line

$$\Delta L = \int_{\nu} \Delta L_\nu(\nu)\mathrm{d}\nu. \qquad (4.100)$$

The line profile is essentially determined by the Doppler width (see Vol. 3. Sect. 7.5). Since the Doppler width depends on the mass of the absorber, the frequency ν_{ki} and the temperature T, it allows the determination of the temperature at the location of the absorbing atom, since all other parameters are known.

Often an equivalent spectral width ΔA_λ is defined, which has a rectangular profile with complete absorption and an area that represents the same total absorption as the Fraunhofer line (Fig. 4.115).

With telescopes of sufficient spatial resolution and in particular with space telescopes, where the air turbulence of the Earth atmosphere is eliminated, one finds that the photosphere is by no means a quiet smooth surface but shows a granular structure that changes in time and reminds one of a boiling liquid surface (Fig. 4.116).

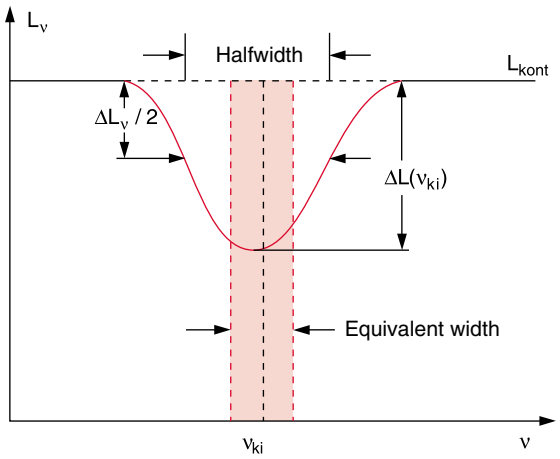


Fig. 4.115 Halfwidth and equivalent width of a spectral line

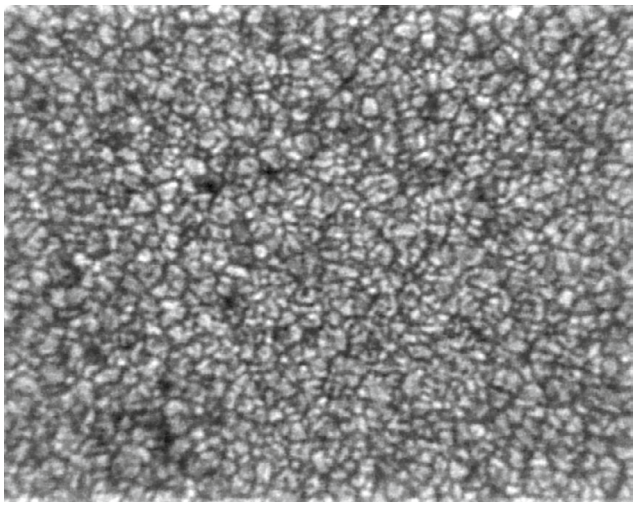


Fig. 4.116 Section of the Sun surface showing the granulation (picture size: 30.000 km × 30.000 km) [32]

It consists of an irregular structure of bright small spots, which are separated by darker channels (Figs. 4.116 and 4.117). The mean diameter of these granules is about 1000 km. The brightness difference between dark and bright spots is about 15%, which gives according to the Stefan-Boltzmann law a temperature difference of 200 K. Measurements of the Doppler-shifts in these granules prove that they are caused by uprising (bright spots) and falling (dark spots) gas packets, which confirms the analogy with the picture of a boiling liquid. The radial velocity of these gas packets can be up to 1 km/s.

4.6.8 Chromosphere and Corona

Above the photosphere, which is the major source of the Sun radiation, there are further gaseous layers, with

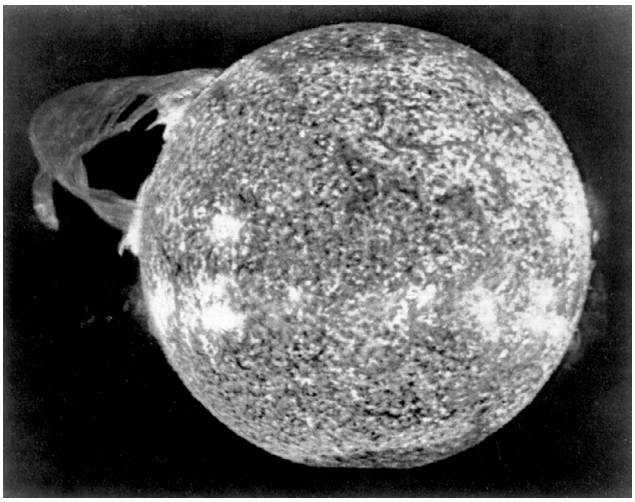


Fig. 4.117 Granulation of the photosphere surface with a protuberance (photo taken by Skylab in the VUV spectral range at $\lambda = 30.4$ nm of the He^+-ion line (NASA))

densities that are much lower than those of the photosphere. Therefore, they cannot be seen at normal Sun observations.

During a total Sun eclipse, however, where the Sun disc is completely occluded, the luminous phenomena of these layers offer a spectacular and impressive picture (Fig. 4.118). The *Chromosphere* extends as a thin layer about 3000 km above the photosphere where it changes over into the much more extended *Corona*, which can reach more than 1 million km above the edge of the Sun disc (Figs. 4.119 and 4.120).

Today these layers can be observed also without Sun eclipses with modern Coronagraphs [33].

There are several specially designed Sun telescopes (Fig. 4.120) for the observation of time-varying phenomena on the Sun surface (Fig. 4.121).

In these telescopes, the image of the Sun in the image plane is concealed by a blackened disc. Scattered light is eliminated by apertures. Therefore, only the corona is imaged into the observation plane. Through narrow spectral interference line filters, specific emission lines of the corona can be selectively detected. The observed spectrum is dominated by the red H_α-line of the hydrogen atom and therefore the corona has a reddish appearance.

A special design of a coronagraph was developed by Bernard Lyot where the occulting disc was replaced by a prism, which reflects unwanted light rays out of the image plane (Fig. 4.122).

Remark Coronagraphs are more and more used for the detection of extra-solar planets, where the image of the much brighter central star is blocked in the Coronagraph.

The density of the Chromosphere decreases exponentially with the height h, following the barometric heights

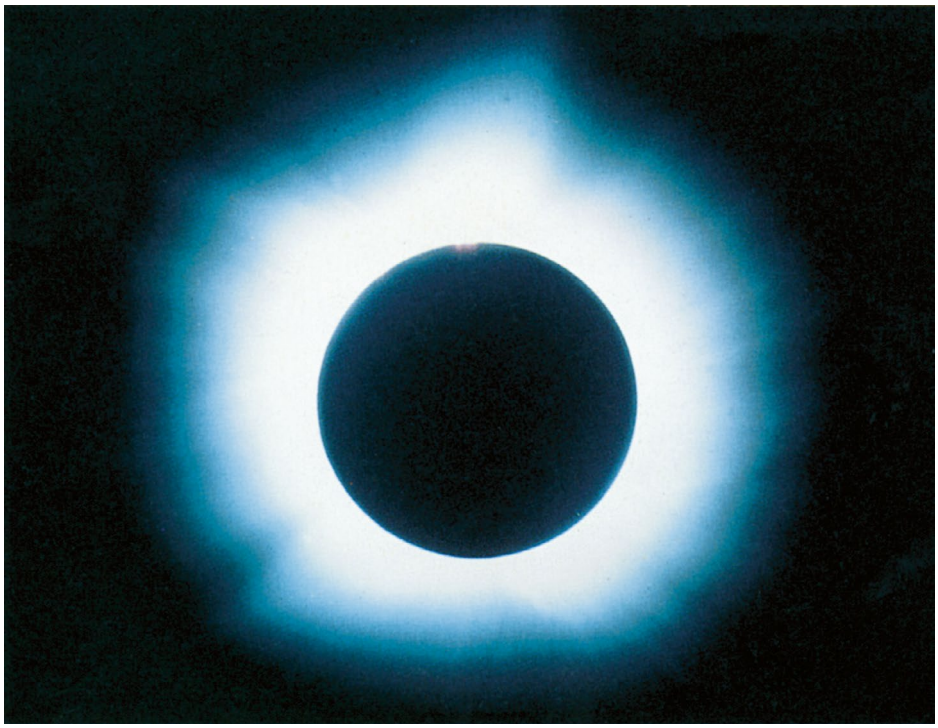


Fig. 4.118 Total Sun eclipse. The photo was taken at the 7th of March 1970 during a maximum of the sunspot number. The obscured photosphere is surrounded by the Sun corona

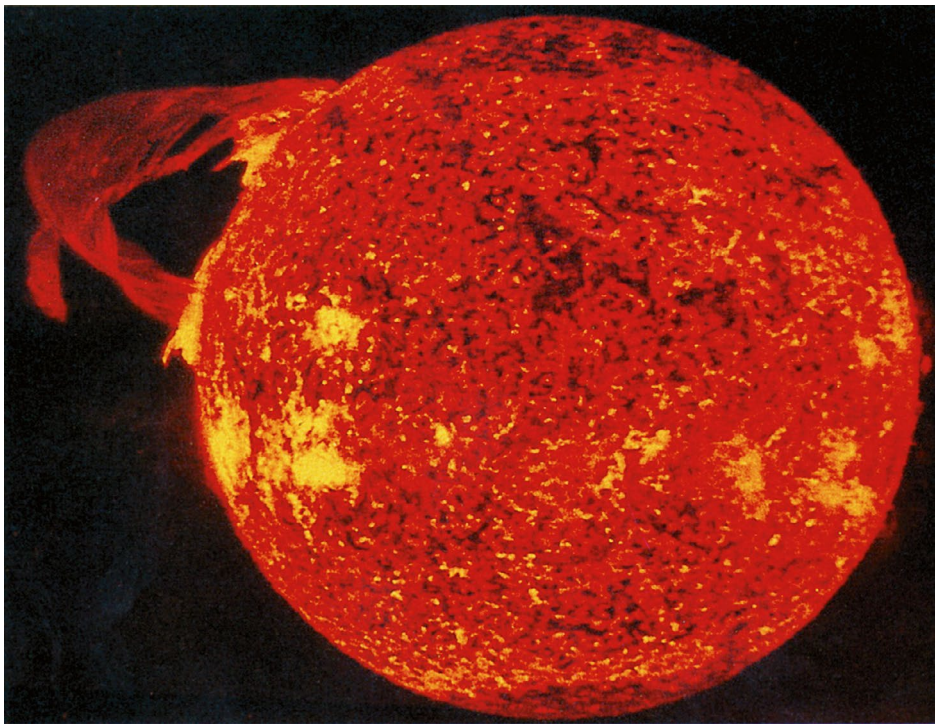


Fig. 4.119 Sun granulation and protuberance photographed at the 21st of August 1993 in the UV-Spectral range. The protuberance extends to 600.000 km above the Sun surface (NASA)

formula. The temperature, however, increases from a value $T_1 = 4500\,\mathrm{K}$ at the outer edge of the photosphere to $T_2 = 10.000\,\mathrm{K}$ at the transition to the corona, where it further increases up to $10^6\,\mathrm{K}$ in the Corona (Fig. 4.123).

The emission of the Corona shows a pronounced line spectrum, contrary to the spectrum of the photosphere. Up to now, more than 3000 lines could be identified. The strongest lines correspond to transitions in the H-atom, the

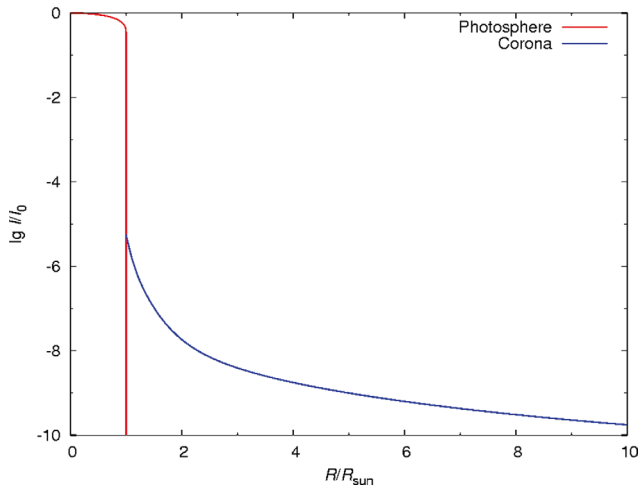


Fig. 4.120 Logarithmic brightness profile of Photosphere (red) and Corona (blue) (https://commons.wikimedia.org/w/index.php?curid=24461)

Fig. 4.121 Richard-B-Dunn solar telescope in the Sacramento mountains

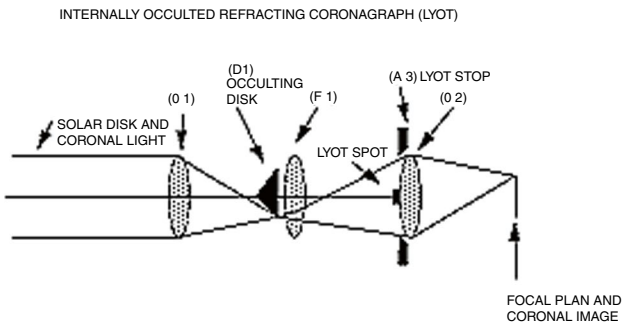


Fig. 4.122 Schematic drawing of the Lyot-type Coronograph

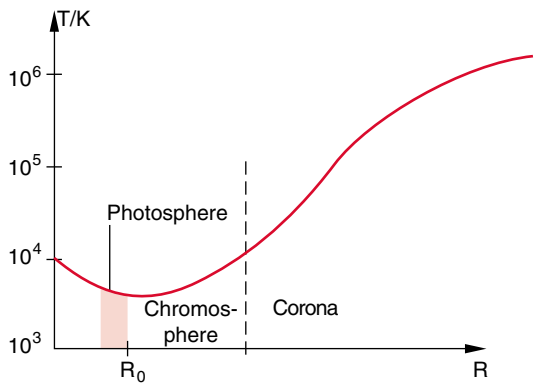


Fig. 4.123 Temperature progression in the Sun atmosphere

He-atom and the He^+-ion. Furthermore, some lines of metals were identified, such as Mg, Fe, Mn or Na. The fact, that no spectral continuum is present, proves that there are no H^--ions due to the high temperature. The gas density is so low that in spite of the high temperature the total heat energy content of the corona is small. The particles at the upper edge of the Corona show a Maxwellian energy distribution. Particles at the high energy tail of this distribution have enough energy to leave the Corona and fly away into the interplanetary space as *solar wind*. The mass loss of the Sun due to this solar wind is about $10^{-13} M_\odot$ per year. It is therefore much smaller than the mass loss due to the emitted Sun radiation. Close to the Earth the flux of particles in the solar wind amounts to $(5-10) \cdot 10^4$ particles/$\left(\mathrm{m}^2\ \mathrm{s}\right)$.

The experimental fact that the temperature of the corona is astonishingly higher than that of photosphere and Chromosphere was for a long time an unsolved puzzle because there is apparently no additional energy source in the Corona. The up to now most plausible models of heating mechanisms assume that there are shockwaves, generated by turbulences of the uprising gas packets in the photosphere, which propagate into the Corona. An additional heating mechanism could be magneto-acoustic waves of ions, which are generated by time-varying magnetic fields in the Corona, which can transport energy and deposit this energy by collisions in the corona.

The shockwaves which heat up the chromosphere represent pressure waves (sound waves) with a velocity dispersion, where the velocity decreases with increasing heights. The amplitude, however, increases. The waves with higher velocities behind those with smaller velocities overtake the preceding slower waves. This superposition results in an increase of the amplitude and generates shock waves which result in a high temperature in the shockwave front. This increase of the temperature must lead to

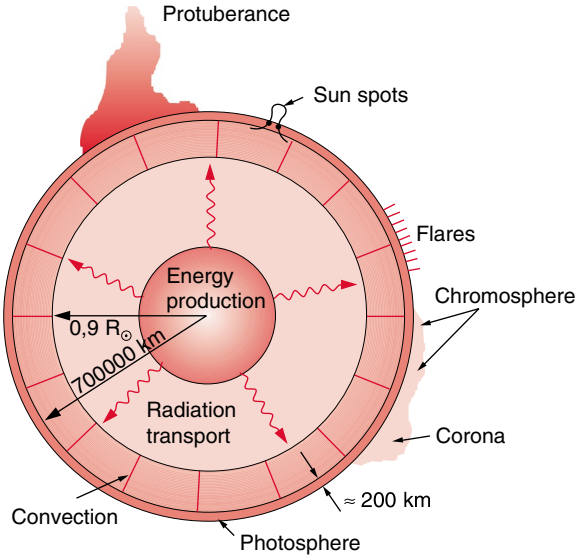


Fig. 4.124 Schematic condensed illustration of the energy production and energy transport in the Sun and the Sun atmosphere, consisting of photosphere, chromosphere and corona

a local decrease of the particle density, according to the equation of state

$$p = n \cdot k \cdot T$$

because the pressure p has to be equal to the gravitational pressure at hydrodynamic equilibrium.

The first experimental proof of the existence of waves was provided by the satellite SOHO (**S**olar and **H**eliospheric **O**bservatory) with three different instruments on board which measured the vacuum ultraviolet radiation of the Sun and its Doppler-shift of spectral lines. It could measure with a coronagraph the Sun corona any time (i.e. not only during a Sun eclipse).

At the end of this section, Fig. 4.124 comprises the model of the "quiet" Sun, accepted today: The energy production by nuclear fusion in the inner part of the Sun, the transport of this energy by radiative transfer for $0 < r < 0.84 R_\odot$ and by convection for $0.84 R_\odot < r < R_\odot$.

4.7 The Active Sun

The Sun shows some phenomena, which change in time and which have their origin below the Sun surface. These phenomena are the **Sunspots**, the **super-granulation**, **facula**, **flares** and **protuberances**. They are all related to variations in time of local magnetic fields. In this section, we will briefly discuss such dynamic processes which lead to observable phenomena at the surface of the Sun.

4.7.1 Sunspots

When observing the enlarged image of the Sun in the image plane of a telescope, some dark spots are visible on the bright disc of the Sun (Fig. 4.125). They show a dark central part (called the Umbra) and a less dark fibrous fringe (penumbra), which shows a filamentary structure. This indicates turbulent currents of ingoing and outgoing hot gas.

The diameter of the umbra is for medium size sunspots about 10.000 km. The lifetime of the sunspots depends on their size and ranges from a few days for small sunspots up to 100 days (about 4 rotational periods of the Sun) for large spots.

Sunspots appear often in groups (Fig. 4.125b), The surface temperature in the umbra is about 1500 K below the temperature of the unperturbed Sun surface.

Since the radiation energy is proportional to T^4 it is lower in the umbra by the factor

$$a = \left(\frac{4300}{5800}\right)^4 \approx 0.3 = 30\% \tag{4.101}$$

compared to the luminosity outside the sunspot.

In order to define abundance of sunspots in a quantitative way, the *relative sunspot number* is defined as

$$R = C \cdot (S + 10 \cdot G) \tag{4.102}$$

where C is a constant depending on the observation instrument. G is the number of sunspot groups and S the number of sunspots in a group.

Detailed observations have shown, that between the quantity R, the area F covered by sunspots and the total area $A = \pi R_\odot^2$ of the Sun disc the relation

$$F = 16.7 \cdot 10^{-6} \cdot A \cdot R \tag{4.103}$$

holds.

▶ The relative number R is therefore a statistical measure for that fraction of the Sun surface, that is covered by sunspots.

Long-term observations showed that the sunspot-relative number R exhibits a periodic variation with a mean period of 11 years (Fig. 4.126) (sunspot cycle). The time period fluctuates randomly between 9 and 13 years.

Observing the sunspots for several days, one notes that they wander from west to east over the Sun disc. This indicates that the Sun must rotate. The path of the sunspots proves that the rotation axis is not vertical to the ecliptic but is inclined by the angle $i = 7°15'$ against the normal to the ecliptic.

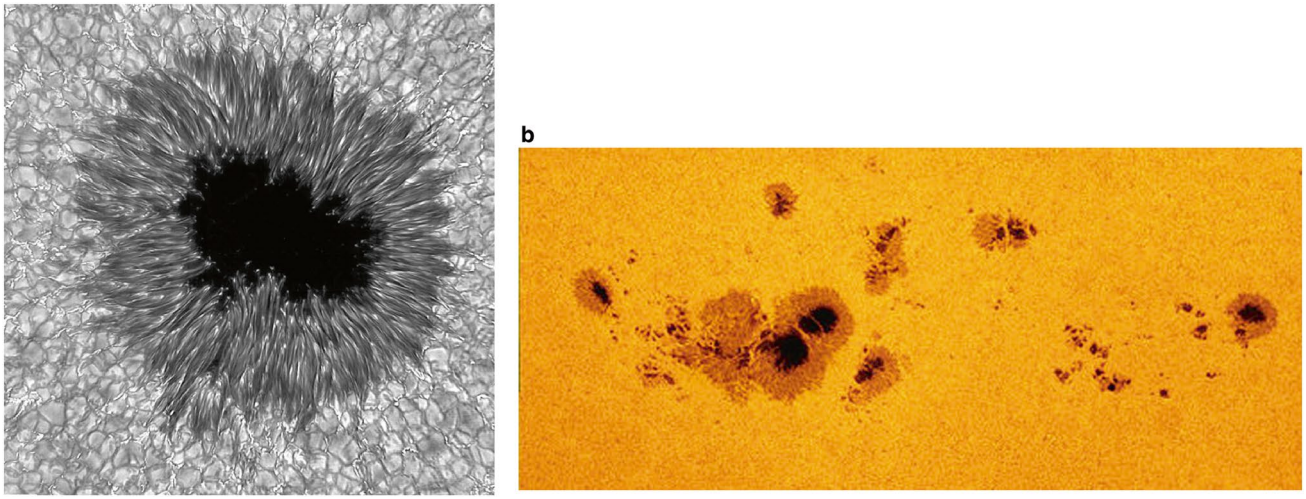


Fig. 4.125 **a** Sunspot with dark umbra and less dark penumbra. **b** A group of sunspots (side length of picture 350.000 km) (NASA)

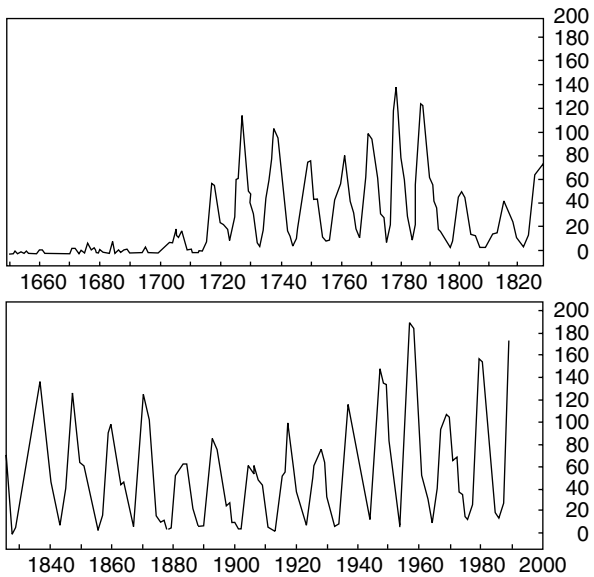


Fig. 4.126 Nearly periodic temporal progression of the sunspot relative number R during the last 240 years

The rotation of the Sun has the same direction as the orbital revolution of the Earth. One must distinguish between the sidereal (relative to a fixed star) and the synodical (relative to the Earth position) revolution period. It is (see Eq. (4.21))

$$\frac{1}{T_{\odot}^{\mathrm{syn}}} = \frac{1}{T_{\odot}^{\mathrm{sid}}} - \frac{1}{T_{\mathrm{E}}^{\mathrm{sid}}} \tag{4.104}$$

with

$$T_{\mathrm{E}}^{\mathrm{sid}} = 365.256\,\mathrm{d}.$$

It is therefore possible to calculate the sidereal rotation period $T_{\odot}^{\mathrm{sid}}$ of the Sun from the observed synodical period $T_{\odot}^{\mathrm{syn}}$.

For the classification of the position of the sunspots, a coordinate system is defined on the Sun, analog to that on the Earth. The meridians connect north- and south-pole of the Sun as breakthrough point of the rotation axis. The circles of latitude are parallel to the Sun equator.

The heliographic latitude φ starts with $\varphi = 0$ at the Sun equator and has the values $\varphi = +90°$ at the north pole and $\varphi = -90°$ at the south pole. For the determination of the heliographic meridian, the angle L is defined as the angle between the zero meridian and the meridian L. According to an international agreement, the meridian L = 0 is defined as the central meridian at the date 1.1.1854, 12 o'clock universal time.

▶ The observation of the sunspots reveals that the Sun does not rotate as rigid body but shows a differential rotation: At the Sun equator, it rotates faster than at the poles (Fig. 4.127).

This differential rotation can be approximately described by the angular velocity

$$\omega(\varphi) = A + B \cdot \sin^2\varphi \tag{4.105}$$

with $A = 14.18°/d$ and $B = -2.0°/d$

In Fig. 4.127b, it can be seen that the outer layers of the Sun ($r > 0.7\ R_{\odot}$) rotate faster than the interior for $\varphi < 40^{\circ}$ but slower for $\varphi > 40^{\circ}$. This again illustrates that the Sun does not rotate as a rigid body.

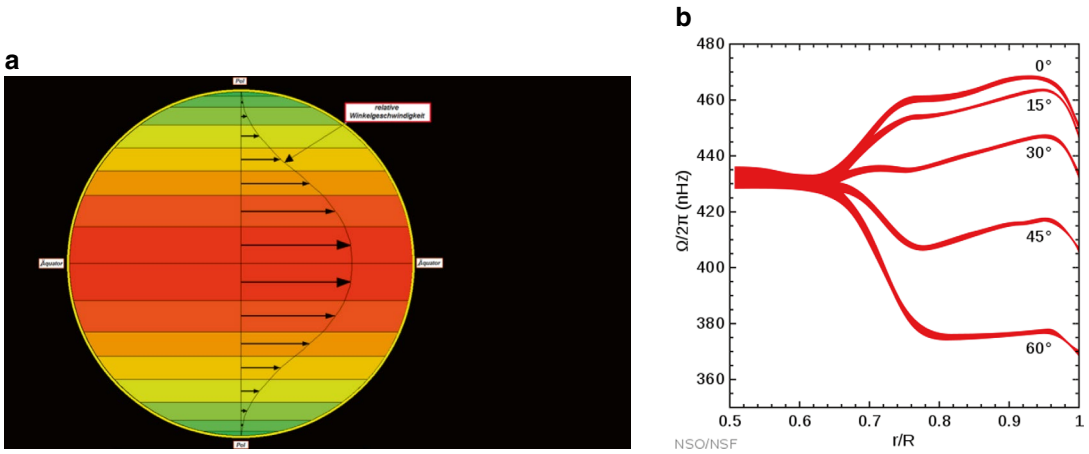


Fig. 4.127 **a** Differential rotation of the Sun. **b** Angular velocity Ω(r) as a function of the distance r from the Sun center for different latitudes (Wikipedia)

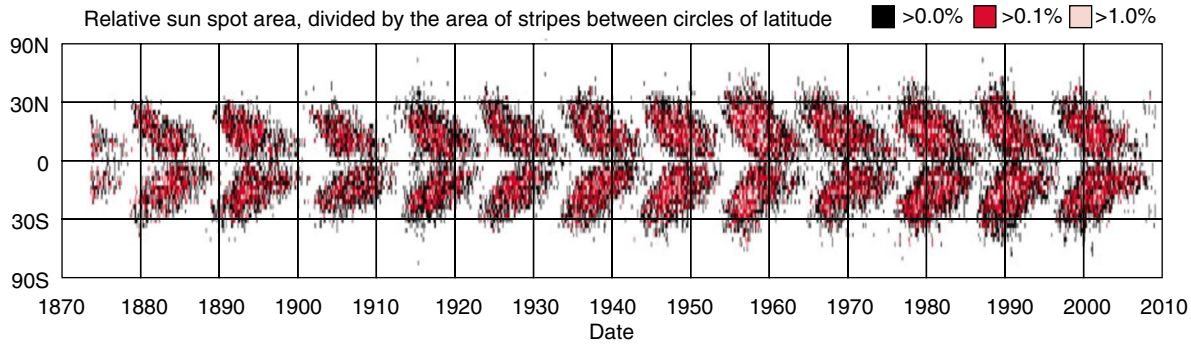


Fig. 4.128 Periodic wandering of sunspots towards the Sun equator. The color code gives the fraction of the Sun disc covered by sunspots (Wikipedia)

The sunspots originate in the northern as well as in the southern heliosphere in medium latitudes around $\varphi = \pm 30°$. During their migration from west to east, they drift towards the equator (Fig. 4.128).

Based on observations of the Zeeman splittings of spectral lines, it could be confirmed that in the sunspots strong radial magnetic fields exist (Fig. 4.129) with field strengths up to B = 0.45T. In the umbra, they are directed either upwards (magnetic north pole) or downwards (magnetic south pole). Most sunspots appear as bipolar groups where the two largest spots have opposite polarity.

These two spots move behind each other on the same latitude. The preceding spot is called p-spot, the following f-spot.

During one sunspot cycle, the polarity sequence is retained. For instance, during one cycle, all p-spots on the northern heliosphere are north poles, all f-spots are south poles, while on the southern heliosphere, the sequence is inverted. During the next sunspot cycle, the conditions are inverted.

All these observations provide the basis of a more realistic model of sunspots and the magnetic field of the Sun.

The observation and detailed investigation of sunspots by Galilei was a great surprise to the Christian contemporaries of Galilei because they had regarded the sun as paradigm of an immaculate pure creation of God which should have no faults.

4.7.2 The Magnetic Field of the Sun

Opposite to the Earth, the Sun has no global magnetic dipole field but strong local fields. The superposition of these local field results in a weak global magnetic field of the Sun. These local magnetic fields varying from place to place can be experimentally studied by measuring the Zeeman splittings of Fraunhofer lines with magnetographs. A magnetograph is the combination of a spectrograph and polarizers (Fig. 4.130). The circularly polarized Zeeman components σ^+- and σ^- are converted by a $\lambda/4$-plate into

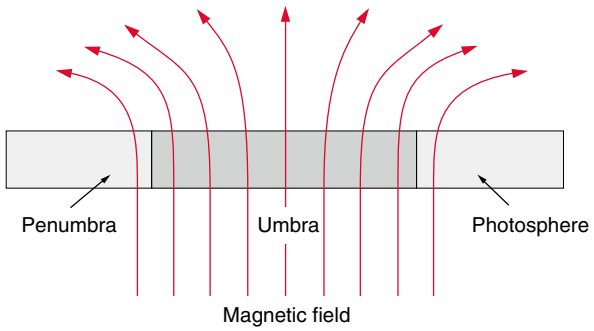


Fig. 4.129 Magnetic field lines in a sunspot

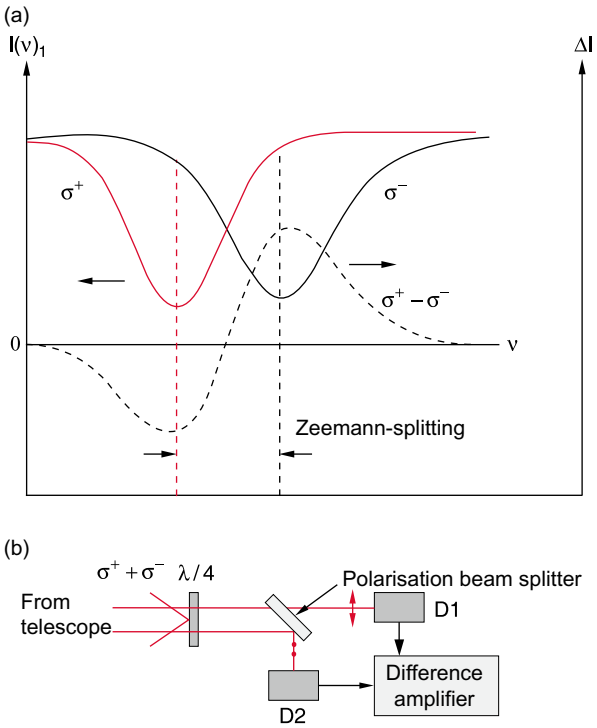


Fig. 4.130 Measuring principle of a magnetograph. **a** Zeeman splitting of a Fraunhofer line. **b** Experimental setup

two linear polarized components, which have mutually perpendicular direction so the polarization vector. They are sent through a polarization beam splitter onto two detectors D_1 and D_2. The measurement proceeds as follows:

The profile of a spectral line is measured by tuning the spectrograph over the spectral range of interest. The transmitted intensity is sent to the beam splitter and the difference of the signals from the two detectors D_1 and D_2 is recorded.

Often the iron line at $\lambda = 525.02\,\text{nm}$ is chosen because the corresponding transition has a large Lande-factor $g = 3$ and therefore the Zeeman splitting is large. The spatial resolution of the local source on the Sun, which emits the radiation, is about 100 km, if measured by a spectrograph on a satellite.

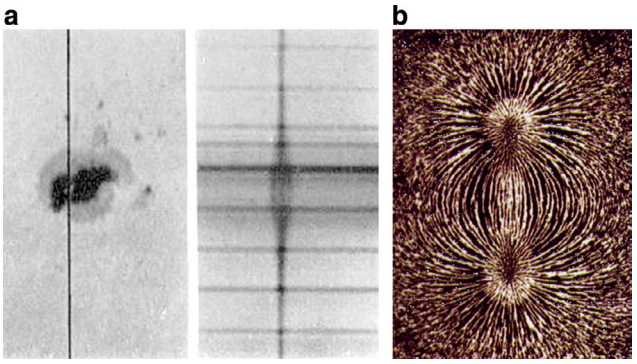


Fig. 4.131 **a** Zeeman splitting of a spectral line in a sunspot (by G. E. Hale, F. Ellerman, S. B. Nicholson, and A. H. Joy). **b** Magnetic field lines in a sunspot pair (NASA)

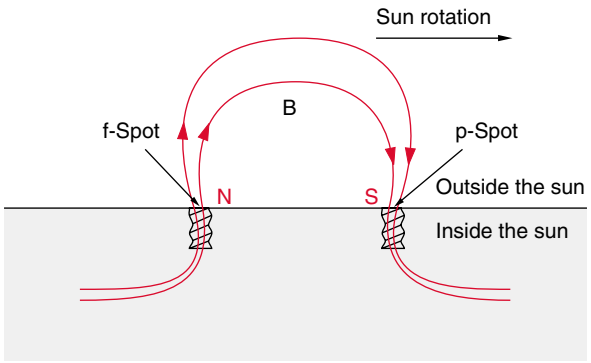


Fig. 4.132 Penetration of the magnetic field lines of a sunspot pair through the Sun surface

In Fig. 4.131a, the Zeeman splitting of a spectral line in a sunspot is shown, and Fig. 4.131b shows the magnetic field lines of a pair of sunspots (dipole pair), obtained by many locally resolved Zeeman splittings. At the sun surface appears a magnetic field line loop with north- and south pole (Fig. 4.132).

In this model, the magnetic field in the sunspots hinders the uprising of hot plasma packets due to the Lorentz force. Therefore, the temperature in the sunspots is lower than that of the unperturbed surroundings. The sunspots correspond in this model to the intersection points of the magnetic field lines, deformed by the uprising plasma packets.

The surface of the Sun is interspersed by magnetic flux tubes, which include magnetic fields up to 0.2 T. They are produced by local convection of hot plasma packets below and above the photosphere. The resulting plasma currents contain ions and electrons with different drift velocities. This causes an electric net current which induces a magnetic field. In the unperturbed photosphere only thin (100–200 km diameter) magnetic flux tubes exist., Since there are uprising and falling plasma packets, magnetic fields of both

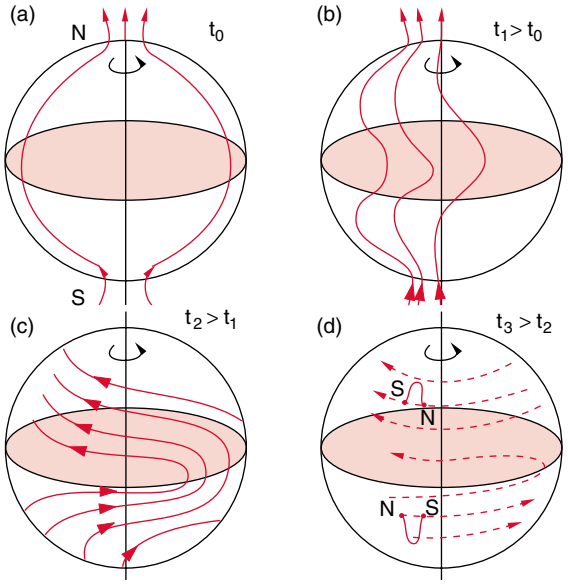


Fig. 4.133 Schematic illustration of the deformation of the magnetic field lines due to the differential rotation of the Sun after the model of Babcock

polarities are present. Since charged particles can only move freely along the magnetic field lines while the Lorentz-force acts on particles moving across the magnetic field lines, the charged particles driven by convection, can carry along the magnetic field and thus deform it. Altogether a global weak magnetic field originates with field lines that proceed in the unperturbed photosphere just below the Sun surface from the south pole to the north pole (Fig. 4.133a).

Due to the differential rotation of the Sun, the entrainment of the magnetic field lines is faster at the Sun equator than towards the poles. This causes a characteristic deformation of the total magnetic field lines (Fig. 4.133b, c), which finally become narrow loops parallel to the circles of latitude (Fig. 4.133d).

4.7.3 Facula, Flares and Protuberances

When observing the Sun disc through a narrow band spectral filter, which transmits only a single strong Fraunhofer line. One sees in the surroundings of a sunspot a network of bright structures, which are due to gas eruptions in the chromosphere (chromospheric facula) (Fig. 4.134), as can be seen when observing facula at the edge of the Sun disc. These facula consist of clouds of bright gas and show an increased Sun activity with magnetic perturbances. They often occur just before a sunspot appears and they remain also after the sunspot has already disappeared.

Larger gas eruptions, which reach far into the corona and which last for several minutes up to hours, are called **flares** (Fig. 4.135). In the flares, electrons and protons are accelerated to high energies by the temporally changing magnetic fields. The accelerated electrons reach velocities up to $v = 0.5c$ and emit *bremsstrahlung*, (resp. Synchrotron radiation) with a spectrum ranging from X-rays to radio frequencies. A further source of the radiation are plasma oscillations induced by the varying charge density in the flares. Here, the electrons oscillate against the protons.

The flares are a source for the solar wind because the radial velocities of the fast electrons and protons can exceed the escape velocity and they can leave the gravitational field of the Sun.

Very spectacular phenomena of the active Sun are protuberances (Figs. 4.136 and 4.119) where charged particles are expelled with large velocities out of the photosphere and then spiral around the field lines of the magnetic field of a bipolar sunspot pair. This results in large luminous arcs, which reach far above the photosphere. The comparison with the size of the Earth illustrates the huge extend of such protuberances.

Since the light emission intensity is still far below that of the photosphere, one can see these phenomena only during Sun eclipses or with as coronagraph. On can photograph

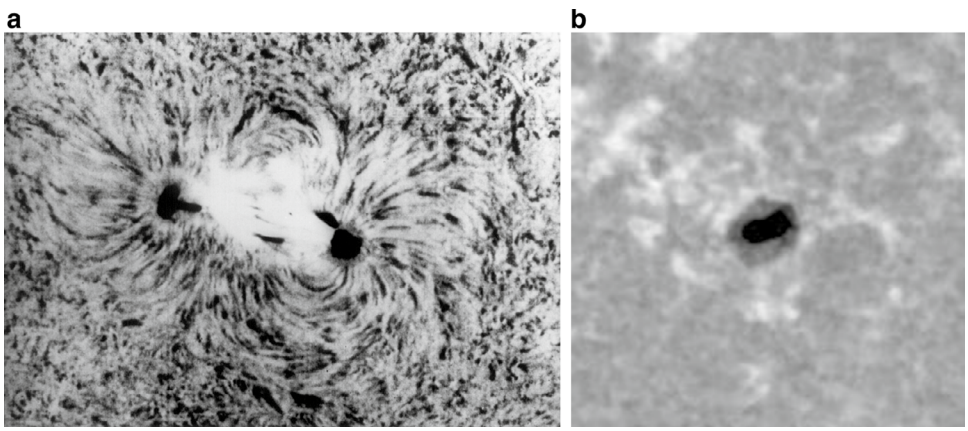


Fig. 4.134 **a** Chromospheric facula in the vicinity of a sunspot pair. **b** An extended facula region (Arbeitsgruppe-Fackeln (vdsastro.de)

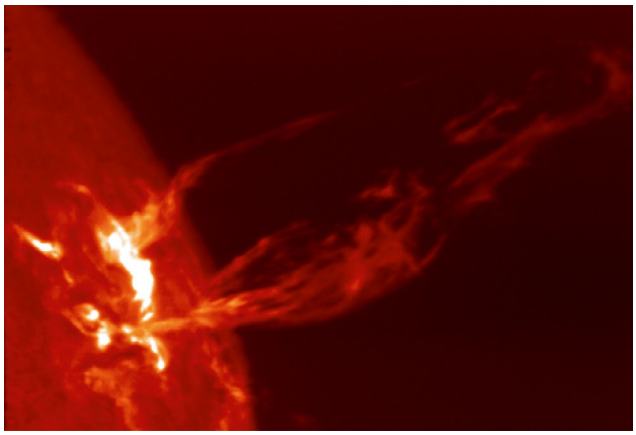


Fig. 4.135 Sun flare (U.S. National Oceanic and Atmospheric Administration)

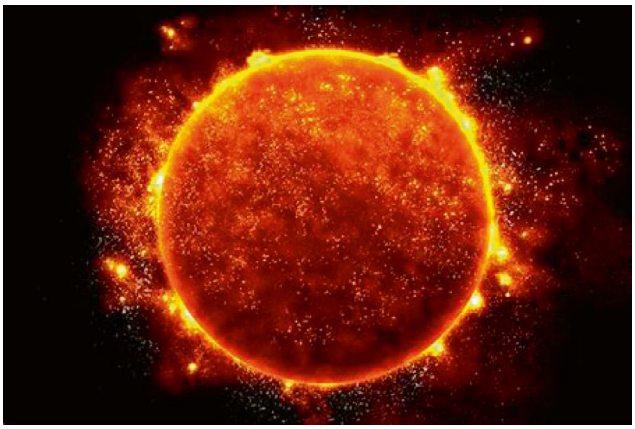


Fig. 4.137 Sun protuberance observed through an UV filter by the satellite SOHO

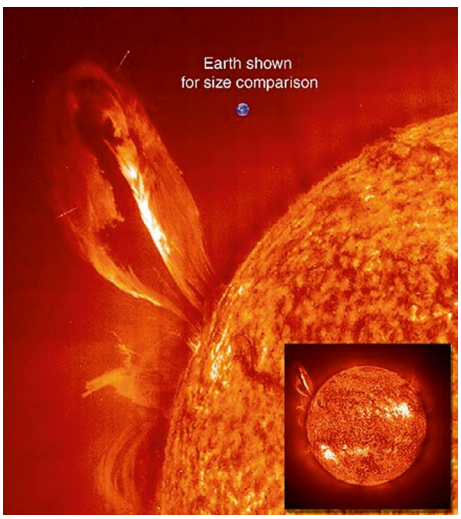


Fig. 4.136 Sun protuberance. For comparison, the Earth is shown as a small blue spot in the upper part of the photograph (Wikipedia)

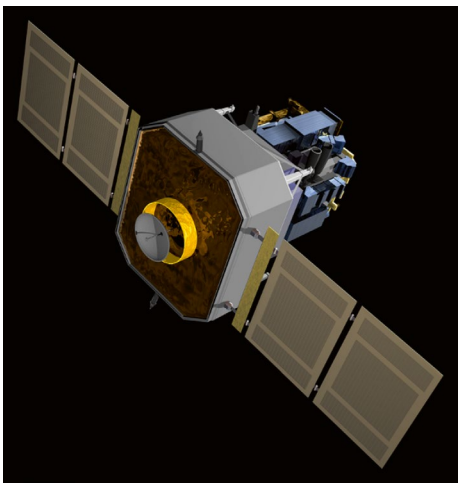


Fig. 4.138 The satellite SOHO (NASA)

them through a narrow band spectral filter which transmits only one prominent line. Flares and protuberances can be recognized extra clearly when observing them in the X-ray or vacuum UV spectral range. Therefore, pictures obtained in the UV from satellites outside the Earth atmosphere yield particular impressive views of protuberances (Fig. 4.137).

One distinguishes between static protuberances that appear as large elongated luminous clouds in the corona and do not change their form for several weeks, and eruptively uprising protuberances, which appear above sunspots as large bright bows and which last only for a short time.

These examples illustrate that the Sun is a very active dynamic gas ball where the hot gas boils, bubbles and whirls. In spite of great progress in the understanding of the Sun, many of these phenomena are still not fully understood. Since 1995 the space satellite SOHO (**So**lar and **H**eliographic **O**bservatory, Fig. 4.138), fixed at the Lagrangian point L_1 about 1.5 Million km away from the Earth, observes continuously the Sun and sends pictures taken at different wavelengths.

4.7.4 The Pulsating Sun; Helio-Seismology

During the last years, detailed time-resolved measurements of the Doppler-shift of spectral lines (e.g. the Fraunhofer lines Fig. 4.113) have revealed that the Sun sphere performs acoustic oscillations. Pictures of the Sun taken by the satellite SOHO have proved that all possible acoustic resonances of a sphere with radius $R_\odot$ and r-dependent sound velocities v(r) indeed occur. The frequencies ν_n of these oscillations depend on the wavelengths λ_n of the eigen-resonances and on the sound velocities v_n. The oscillation frequencies are in the mHz region (Fig. 4.139), the oscillation period

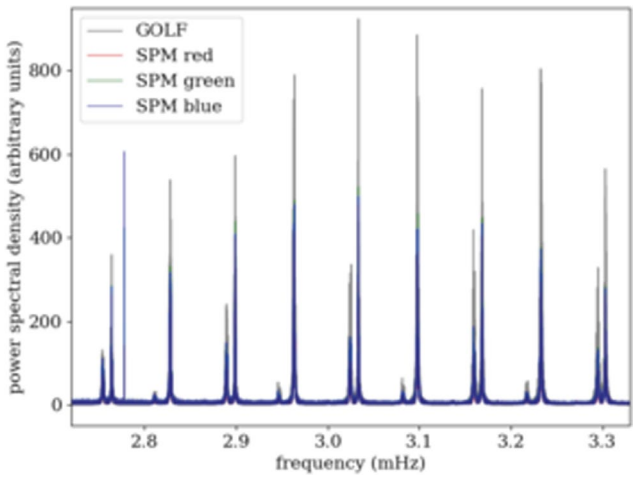


Fig. 4.139 Power spectrum of oscillating seismographic eigen resonances (data taken from instruments on the satellite SOHO, NASA)

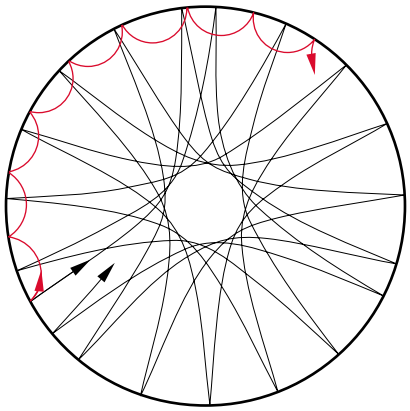


Fig. 4.140 Spatial extension of acoustic resonance waves in the Sun

$T_n = 1/\nu_n$ are several minutes, the oscillation amplitudes range from 10 to 100 m and the wavelengths λ_n vary between 10^3 and 10^5 km. The maximum speed in the maxima of the oscillation reaches 500 m/s.

The great advantage of helio-seismology is the fact the it is possible to look into the interior of the Sun because the resonance oscillations can extend far into the interior of the Sun (Fig. 4.140).

Since frequencies and wavelengths depend on the sound velocity v(r), which itself depends on density and pressure, measurements of frequencies and wavelength of the resonance oscillations give information about the structure of the interior and can confirm Sun models. In Fig. 4.141, the gas currents around a dipolar sunspot in a depth of 1000 km below the Sun surface are shown, indicated by the arrows.

The spectral distribution of the eigenfrequencies (Fig. 4.142) gives information about the frequencies and the intensity of the different modes and therefore also about the mechanism, which induce these oscillations. The velocity

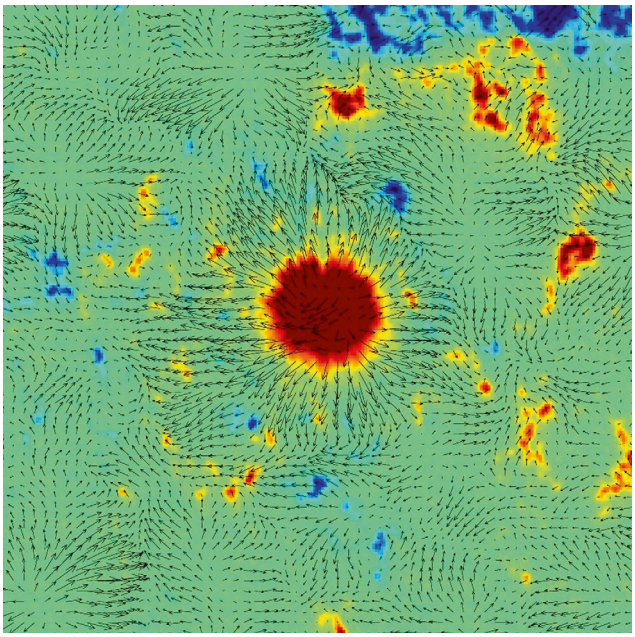


Fig. 4.141 Map of gas currents around a sunspot in a depth of 1000 km represented by the arrows. The side length of the picture is 10.000 km (L. Gizon, MPI für Sonnenforschung, Katlenburg Germany)

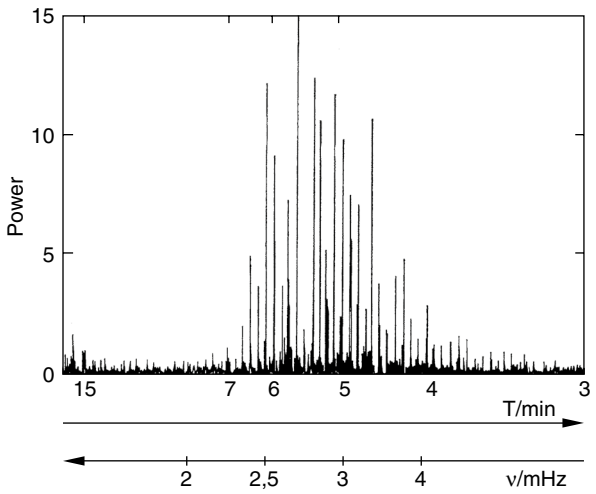


Fig. 4.142 Frequency spectrum of acoustic minute oscillations [29]

increases with increasing depth and therefore also the wavelength of the resonance modes. In Fig. 4.143, the patterns of some P-type eigen resonances are illustrated.

Summary

– Observation of celestial bodies from Earth bound stations is restricted to some spectral windows transmitted through the Earth atmosphere (visible and near infrared

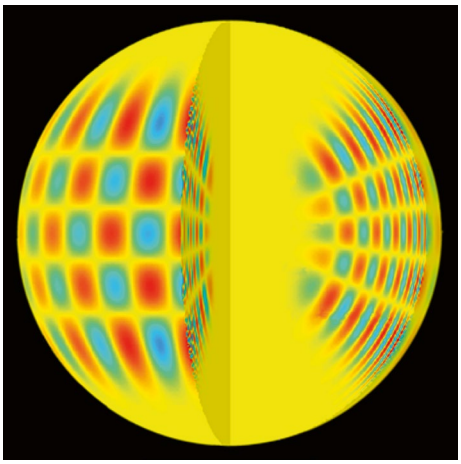


Fig. 4.143 Computer-generated profiles of P-type acoustic eigen resonances of the interior Sun and on its surface (https://physik.cosmos-indirekt.de)

spectrum ($\lambda = 0.4$–$1\ \mu m$), Millimeter region and radiofrequency range. The spectral regions detected by Satellites outside the atmosphere are only restricted by the spectral response of the detectors.

- Our solar system consists of the Sun, eight planets with their moons, dwarf planets, many planetoids, comets, meteorites, dust particles and the interplanetary gas particles.
- The motions of planets, planetoids, comets and moons in central gravitational fields are governed by Kepler's laws.
- Perturbations by neighboring planets or moons cause distortions of the central fields, which result in deviations from the Kepler orbits. Periodic perturbations can even lead to unstable orbits. This occurs particularly for planetoids and the ring systems of the four big planets.
- If the radius vector of an outer planet has the same revolution angle as the Earth, the planet is in opposition if the Earth stands between planet and Sun. It is in conjunction if the Sun stands between Earth and planet. Inner planets can never be in opposition, but they show an additional conjunction point, if the planet is between Sun and Earth.
- During a sidereal revolution period of a planet, the radius vector from the Sun to the planet covers the angle 2π.
- The synodal revolution period is the time between two successive oppositions of the planet. For inner planets, the synodal periods are longer than the sidereal periods, for outer planets they are shorter.
- Mercury and Venus have no moons, the Earth has one, Mars has two very small moons, Jupiter has at least 69 moons, Saturn at least 62, Neptune 14, Uranus 27. The solar system has therefore all together at least 175 Moons.
- The inner (Earth-like) planets have densities of $\varrho = 4 - 5\,\mathrm{kg/dm^3}$and consist mainly of solid crust and liquid magma in their inner part. The outer planets (Jupiter-like) have gas envelopes and in their inner part liquid and solid matter. Their mean density is $\varrho = 0.7 - 1.6\,\mathrm{kg/dm^3}$, which is Sun-like. Their mass is much larger than that of the inner planets.
- The giant planets Jupiter, Mars, Uranus and Neptune have a ring system, which consist of dust particles and mineral rock fragments. It is most pronounced for Saturn. It is located in the equatorial plane of the planets. Its thickness is only about $0.1 - 0.2$ km, its radial extension, however, reaches up to 500.000 km. The ring structure can be explained by resonance of the periods of ring particles and moons of the planet.
- Planetoids are small celestial bodies, moving on orbits around the Sun between the orbits of Mars and Jupiter. Their orbits can become unstable if resonances occur between their revolution period and that of Jupiter.
- Comets are small celestial bodies consisting of ice, dust and a mineral core. Their origin is probably located in the Oort's cloud, an outermost region of the solar system. Their orbits are changed due to gravitational perturbations by nearby passing stars into largely elongated ellipses around the Sun. At the closest approach to the Sun part of the surface of comets evaporates which forms the coma and the extended tail of the comet.
- Meteors (shooting stars) are luminous spurs at the night sky, caused by microparticles from comet tails, which are crossed by the Earth on its orbit around the Sun.
- Meteorites are mineral rocks of medium to large size, which mainly originate from planetoid fragments. On their way close to the Earth, they are attracted and fall onto the Earth surface. On their path through the Earth atmosphere, they are heated up and can be seen as brilliant fire balls.
- The Sun gets its radiation energy from nuclear fusion in the central part. The energy produced is transported to the surface mainly by radiative transfer up to $r = 0.84 R_\odot$. In the range $r > 0.84 R_\odot$, convection takes over. The main transport time between the production of the energy and its radiation from the surface is about 10^7 years.
- The effective surface temperature of the Sun is 5800 K. The temperature rises with decreasing distance r from the center. The central temperature is $15 \cdot 10^6$ K.
- The continuous spectrum of the Sun radiation is emitted from the photosphere, a thin outer layer with about 200 km thickness. The spectrum is generated by the recombination $H + e^- \rightarrow H^- + h \cdot \nu$ of electrons with neutral hydrogen, forming negative ions.

- Above the photosphere extended luminous plasma regions exist (chromosphere and corona) which have a low density but high temperatures. They can be only seen during Sun eclipses or with coronagraphs, where the much brighter light of the photosphere is masked.
- The Sun shows a differential rotation. The sidereal rotational period increases from $T_{\odot}^{\mathrm{sid}}(\varphi = 0) = 25\,\mathrm{d}$ at the Sun equator to $T\left(\varphi = 60^{\circ}\right) = 31\,\mathrm{d}$ at the heliographic latitude of 60°.
- The Sun possesses a weak global magnetic field but strong local magnetic fields.
- The Sun shows many time-dependent dynamic phenomena, such as sunspots, flares, faculae and protuberances. Periodic small changes of the Sun radius by sound waves are explored by the helio-seismology, which gives information about the inner part of the Sun.

Problems

4.1 A comet has in its closest approximation the distance $r_{\min} = 0.5\,\mathrm{AU}$ to the Sun and the velocity $v = 50\,\mathrm{km/s}$. What is its largest distance from the Sun? Calculate the revolution period.

4.2 Derive Eqs. (4.12) and (4.13).
Hint: Multiply (4.2c) vectorial with $\boldsymbol{L} = m(\boldsymbol{r} \times \dot{\boldsymbol{r}})$ and then scalar with $\mathbf{r}$.

4.3 Calculate the apparent luminosity of Mars during the opposition in the year 1982 when the distance Mars-Sun was 1.64 AU. Use the data for the apparent magnitudes $m_1 = -1.6^{\mathrm{m}}$ during the opposition in 1975 when $r_1 = 1.55\,\mathrm{AU}$. How large is the albedo of Mars based on these data?

4.4 A planetoid has the distance $r = 3\,\mathrm{AU}$ from the Sun and the visual luminosity $\mathrm{m_{vis}} = +10.1^{\mathrm{m}}$ at the time of its opposition. What is its diameter if its albedo is 0.15? How accurate can its diameter be determined, if its albedo is known within 30% and its luminosity within 10%?

4.5 How large would be the oblateness of Jupiter if it would consist to 100% of gas?
(a) At a constant density ϱ_0?
(b) For $\varrho(r) = r_0\left(1 - r/R_{\mathrm{Jup}}\right)$?
(c) Calculate the mass of Jupiter from the orbit period $T = 7.1546$ d of its moon Ganymed with a distance $a = 1.0705 \cdot 10^6\,\mathrm{km}$ from the center of Jupiter.

4.6 Calculate the difference of the gravitational forces acting on the outer and inner side of the moon Io as a function of its distance r from Jupiter. At which value of r is the Roche limit?

4.7 Show that the Lagrange points $\mathrm{L_i}$ in Fig. 4.75 are located on a circle around the Sun with the radius $\mathrm{R_{Jup}}$ of Jupiter and that they form an equilateral triangle and represent minima of the potential energy.

4.8 Why can a space probe win energy during the passage through the gravitational field of a planet? Calculate the energy gain resp. energy loss dependent on the relative velocity and the mass of the planet.

4.9 Derive Eq. (4.44).

4.10 Derive Eq. (4.64a) for the temperature gradient in case of radiative energy transport.

4.11 What is the gravitational pressure at the surface of the Sun, if the density ρ(r) from the edge of the photosphere $(r = R_{\odot})$ decreases exponentially as $\varrho(r) = \varrho(R_{\odot}) \cdot \mathrm{e}^{-(r-R_{\odot})/a}$ with $a = 100\,\mathrm{km}$?

4.12 How large is the escape velocity of a particles from the Sun, from Jupiter and from the Earth moon?

4.13 Derive Eq. (4.68).

4.14 How large is the hydrostatic pressure at R/2 of the Sun radius R.
(a) if the density $\varrho(r) = \overline{\varrho}$ is assumed to be constant?
(b) for $\varrho(r) = \varrho_0(1 - r/R)$.

References

1. 1969 Moon Landing—HISTORY
2. Apollo 11—Wikipedia
3. Donald L. Turcotte; Gerald Schubert (25 March 2002). Geodynamics. Cambridge University Press.
4. Earth's energy budget—Wikipedia
5. https://en.wikipedia.org/wiki/Origin_of_the_Moon
6. Analysis of Mars relevant minerals—suggestions for next missions
7. https://en.wikipedia.org/wiki/Solar_system
8. Marquis M., Averyt K.B., Tignor M., Miller H.L. (Hg.) (2007) Climate change 2007: The physical science basis. Contribution of working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, New York: Cambridge University Press, 996 pages, ISBN 9780521705967 (Website)
9. https://de.wikipedia.org/wiki/Venus_%28Planet%29
10. Don L. Anderson, Theory of the Earth, Blackwell (1989) Is a textbook dealing with the Earth's interior and is now available on the web
11. Thomas E. Cravens: Planetary ionospheres and magnetospheres. Pergamon Press, Oxford 1997
12. https://de.wikipedia.org/wiki/Jupiter_(Planet)
13. List of Jupiter Moons—Wikipedia
14. Rings of Jupiter—Wikipedia
15. Rings of Uranus—Wikipedia
16. https://en.wikipedia.org/wiki/Moons_of_Neptune
17. Moons in our solar system. Windows to the Universe https://windows2universe.org/our_solar_system/moons_table.html&edu=high
18. https://theplanets.org/moons
19. (90377) Sedna – Wikipedia
20. https://www.bing.com/search?q=dwarf+planets&form=ANNT
21. Komet—Wikipedia
22. List of Cometes wkipedia
23. https://de.wikipedia.org/wiki/Tschurjumow-Gerassimenko
24. https://www.cliffsnotes.com/study-guides/astronomy/the-sun-a-representative-star/properties-of-the-sun, https://en.wikipedia.org/wiki/Sun

25. https://www.swpc.noaa.gov/sites/default/files/images/u33/Activity_2.pdf
26. GALLEX—Wikipedia
27. https://www.bing.com/search?q=sun+neutrinos&form=ANNTH1&refig=18dd1225ac2b4397b2e0b744020b4ca6
28. https://de.wikipedia.org/wiki/Neutrinooszillation
29. Henny J.G.L.M. Lamers and Emily M. Levesque: Radiative Energy Transport (I OP Publishing Ltd 2017)
30. Photon diffusion—Wikipedia
31. Coronagraph—Wikipedia
32. https://en.wikipedia.org/wiki/Granule_(solar_physics)
33. Sunspot—Wikipedia
34. Ein neuer Blick auf das Magnetfeld der Sonne | Max-Planck-Gesellschaft (mpg.de)
35. Basu, S.; Antia, H.M. (2008). "Helioseismology and Solar Abundances". Physics Reports. 457 (5–6): 217–283
36. Forschungsfelder Helioseismologie und Asteroseismologie | Max Planck Institut für Sonnensystemforschung (mpg.de). http://gong.nso.edu/info/helioseismology.html
37. Helioseismology (stanford.edu)

Birth, Lifetime and Death of Stars

5

Contents

5.1 The Stars in Our Neighbourhood 122
5.2 Direct Measurements of Star Radii 124
5.3 Binaries and the Determination of Masses and Radii of Stars 125
5.4 The Birth of Stars 130
5.5 The Stable Stage on the Main Sequence 136
5.6 Star Development After Leaving the Main Sequence . . 140
5.7 End-Stages of Stars 145
5.8 Quasars 161
5.9 Observable Phenomena During the Final Stage of Stars 161
5.10 Summarizing the Star Development 170
5.11 Thought Provoking 172
Summary 173
Problems 174
References 174

In former times it was generally believed that stars and the whole universe were static objects which do not change in time but are constant forever "*from eternity to eternity*".

Nowadays we know that stars are changing in time and also the universe as a whole modifies. It expands, galaxies are created and might collide with each other, stars are born by contraction of huge gas- and dust clouds. They pass through different development stages, like humans, but on a very much longer time scale. During these stages they emit radiation until they reach their final state, where they go out, either very fast after their explosion or as slowly cooling celestial objects. Star explosions provide material for the formation of new stars. In this sense one can speak of birth, life and death of a star.

The life of a star generally lasts for many millions up to billions of years. Only very hot and massive stars last for less than a million years. Within the observation time of a human generation it is only possible to detect the present actual state of a star. In order to gain information about the development stages of a star one has to observe many stars at the same time but each at different development stages. The fraction of those stars in a certain stage is proportional to the time a star spends in this stage.

This is completely analogue to the following situation: A visitor to a foreign country wants to know about the development stages of the people in this country. He meets during his short visit people at different stages of age: Babies, children, adults and seniors. He divides them into corresponding classes. From the fraction of people in the different classes he concludes about the duration of the different stages of life which is proportional to these fractions. If the population is stationary over a longer period of time, this conclusion is correct. The visitor can obtain a true model of the lifetimes of the different stages of life, even if his visit is only very short compared to the mean lifetime of each individual.

We must at first discuss, which properties of stars are indicators for their lifetimes and for their present state. We must learn, how these properties can be measured and which star models for the star development can be derived from such observations.

The most important properties of a star are its mass. its radius, its luminosity and its spectral type, determined by the spectral intensity distribution of the emitted radiation, which is given by its surface temperature and its chemical composition. We have learned in Sect. 4.8 that the absolute luminosity can be obtained from the measured apparent luminosity, if its distance from the observer is known. Distance measurements and their accuracy therefore belong to the most important problems in Astronomy (Sect. 3.5).

Another method for the determination of the absolute luminosity L is based on the measurement of the star radius R and the surface temperature T. These two quantities allow the calculation of the absolute luminosity, according to the Stefan- Boltzmann relation

$$L = 4\pi R^2 \sigma \cdot T^4 \tag{5.1}$$

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_5

Table 5.1 Number of stars within a volume with radius r around the sun

Radius r around the sun	Number of stars within a volume with radius r
3 pc	12
5 pc	52
10 pc	≈400
30 pc	10^4
100 pc	$3 \cdot 10^5$
300 pc	10^7

where $\sigma = 5.670 \cdot 10^{-8} \mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mathrm{T}^{-4}$ is the Stefan-Boltzmann constant.

We will therefore at first discuss some methods for the determination of star radii.

5.1 The Stars in Our Neighbourhood

The star closest to us is (besides our sun) Proxima Centauris at the distance $r = 4.27$ ly $= 1.3$ pc. The mean star density in our neighbourhood is about 0.1 stars/pc^3. Table 5.1 lists the number of stars inside a sphere with the radius r around the sun. The positions and distances of all stars not only close to the sun but in the whole Milky Way have been measured with an incredible high accuracy by the satellites HIPARCOS and GAIA based on a special trigonometric technique (see Sect. 3.8). Gaia measured more than 2 billion stars in our Milky Way which is more than 2% of all stars.

The distances of about 1000 stars close to the sun could be measured with an error less than 0.1% and that of about 10^6 stars in our Galaxy with less than 1% [1].

This allows the accurate determination of the absolute luminosity L_S from the measured apparent luminosity L_a. Since Gaia also measured the spectral distribution of the star radiation and the temperature of the star surface, it is possible to calculate the star radius R from the relation

$$L_S = 4\pi R_S^2 \sigma \cdot T^4 \Rightarrow R_S = \frac{1}{T^2}\sqrt{\frac{L_S}{4\pi\sigma}}. \qquad (5.2)$$

In order to facilitate the orientation on the sky with thousands of stars observable with the naked eye, and to distinguish between the different stars, astronomers have named already very early certain groups of stars subsumed into **star constellations**. Examples for well-known star constellations are the *Greater Bear* (Ursa Major Fig. 5.1), *Orion*

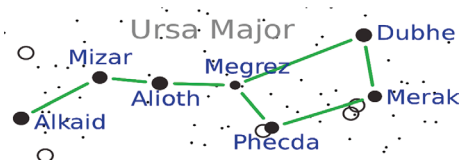


Fig. 5.1 Star constellation ursa major (https://www.constellation-guide.com/constellation)

Fig. 5.2 Star constellation orion (https://www.bing.com/images/search?view=detailV2&cc)

(Fig. 5.2). Auriga, Pegasus, Aries, Bootes, Capricornus. The different stars in such a constellation have generally arabic names because they were first named by arabic astronomers. Nowadays they are identified by Greec letters according to their luminosity. For example the brightest star in Ursa Major is denoted as α-Ursae Majoris (arabic Dubhe), the second brightest star is ß-Ursae Majoris (Merak), etc.

Figure 5.3 shows for illustration the star constellations visible around the celestial equator in the north-winter.

Note: Although the different stars in a star constellation have similar angular coordinates, they have generally very different distances from the earth and are therefore not close together.

The ten brightest stars are listed in Table 5.2.

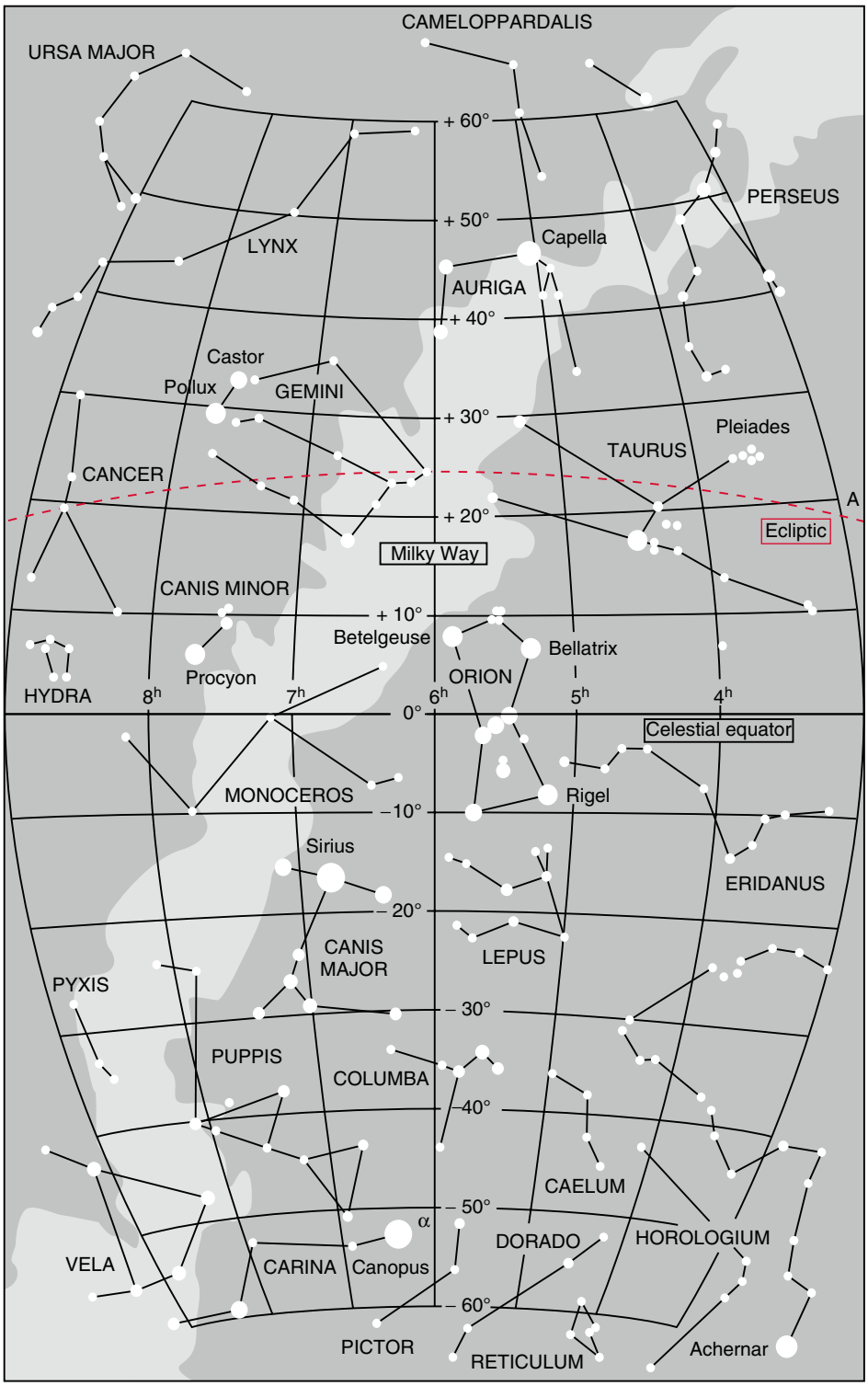


Fig. 5.3 The winter star constellations around the celestial equator [2]

Table 5.2 The ten brightest stars on the sky with their apparent magnitude and their distances

Name	Constellation	Magnitude	Distance/ Ly
Sirius	Canis Major	−1.46	9
Canopus	Carina	−0.72	74
Arcturus	Bootes	−0.04	34
Rigel	Centaurus	−0.01	4
Vega	Lyra	+0.03	25
Capella	Auriga	+0.08	43
Rigel	Orion	+0.12	1400
Procyon	Canis Minor	+0.38	11
https://stellardiscovery.com/brightest-stars-in-the-sky			

Fig. 5.4 Occultation of mars by the moon, just before mars disappears behind the moon. (Andrew Chaikin, Wikipedia)

5.2 Direct Measurements of Star Radii

The angular diameters of stars are even for the closest stars by far too small that they can be measured with conventional trigonometric methods. However, the astrometric satellite Gaia is able to measure the radii of millions of stars with special trigonometric techniques. Therefore some of the older methods for measuring star radii are now obsolete. They will be, nevertheless, shortly discussed here in order to demonstrate the ingenuity of former astronomers.

Example

The closest star (besides the sun) is α-Centauri at a distance $r = 1.34 \cdot pc = 4 \cdot 10^{16}$ m. It has a radius von $R = 8 \cdot 10^8$ m. Its angular diameter, as seen from the earth is $\Delta\alpha = 2R/r \approx 0.01'$. This is by 2 orders of magnitude smaller than the angular resolution limited by the air turbulence (seeing). Even with adaptive optics the diffraction limited resolution of a 1 m mirror telescope is with $\Delta\alpha = 0.1''$ still smaller by a factor of 10. ◄

Each star therefore appears even for large telescopes as the diffraction limited image of a point like source.

In spite of this limitation there are some methods to measure the diameters of close large stars.

One of these methods is the measurement of the occultation time of a star by the moon which is applicable, if the moon crosses the line of sight to a star (Fig. 5.4).

The occultation does not occur uniformly because of the diffraction of the star light at the edges of the moon. The observed diffraction maxima and minima have amplitudes which depend on the diameter of the star. For point like stars the amplitudes are higher than for stars with a finite diameter. Applying the diffraction theory to the photometrically recorded diffraction pattern, the star diameter can be calculated.

A more accurate method uses the periodic occultation of one star by the other in binary systems (see Sect. 5.1.2).

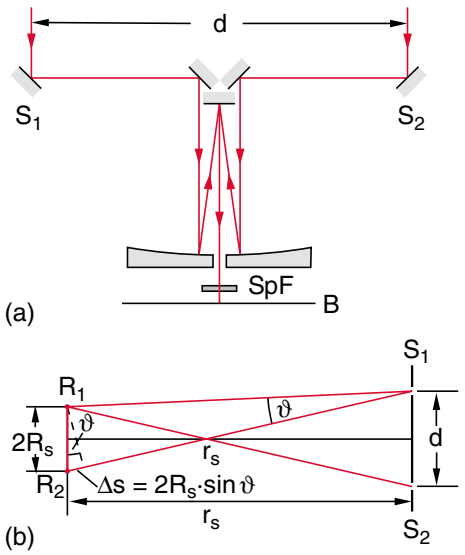


Fig. 5.5 Principle of Michelson's stellar interferometer **a** Experimental setup **b** Conditions for the interference structure

For large stars in our vicinity the star radius can be interferometrically measured with a technique developed by Michelson.

Its principle can be explained as follows (Fig. 5.5). The light from a star falls on the two mirrors S1 and S2, which are separated by the distance d, and is then guided to the entrance of a telescope (Fig. 5.5a). The spectrally filtered radiation intensity is now measured as a function of the distance d. The light from the star with radius R produces in the observation plane an interference pattern, completely analogue to Young's double slit experiment (Vol. 2, Sect. 10.3.2).

With the distance $r_S \gg d$ to the star the maximum path difference for light from opposite edges R_1 and R_2 of the star is (Fig. 5.5b)

$$\Delta s_{\max} = \overline{R_2 S_1} - \overline{R_1 S_1} = \overline{R_1 S_2} - \overline{R_1 S_1} \approx 2R_S \cdot \sin\vartheta = 2R_S \cdot d/2r_S. \tag{5.3}$$

For $\Delta s_{\max} > \lambda/2$ the phase differences

$$\Delta\varphi = (2\pi/\lambda)\Delta s$$

of the waves coming from different parts of the star will in the detector plane randomly fluctuate between $\Delta\varphi = 0$ to $\Delta\varphi = \pi$. This washes out the interference structure in the detector plane B and cannot be analysed.

The condition for the coherent (i.e. phase correlated) illumination of the two mirrors M_1 and M_2 is then

$$\Delta s_{\max} \approx \frac{Rs \cdot d}{r_S} \leq \lambda/2 \Rightarrow R_S \leq r_S \cdot \frac{\lambda}{2d}. \tag{5.4}$$

The contrast of the interference pattern

$$K = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{5.5}$$

is measured as a function of the distance d between the two mirrors (Fig. 5.6). The radius R_s of the star can then be determined from the distance d_c where the contrast becomes K = 0, because

$$R_S = r_S \cdot \frac{\lambda}{2d_S} \Rightarrow d_S = \frac{\lambda}{\Delta\alpha} \tag{5.6}$$

where $\Delta\alpha = 2R_S/r_S$ is the angle, under which the diameter $2R_S$ of the star appears for the observer. If the distance r_S of the star is known, the diameter of the star can be obtained.

The first star, for which the angular diameter $\Delta\alpha = 0.047''$ has been measured, was the red

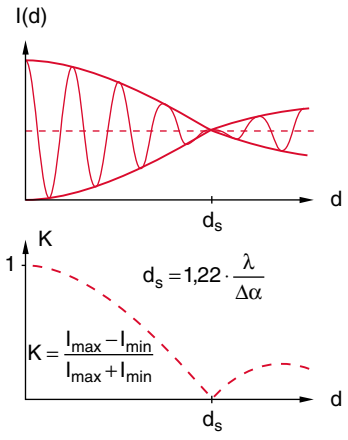


Fig. 5.6 Intensity distribution and contrast in a stellar interferometer as a function of the mirror distance d

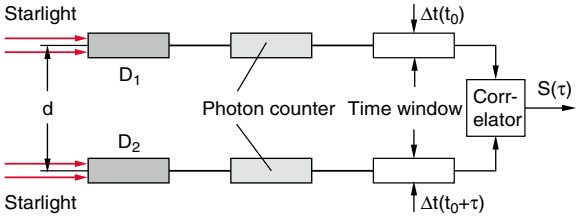


Fig. 5.7 Principle of the Hanbury-Brown-Twiss intensity correlator

giant star Betelgeuse (α-Orionis). With the known distance $r_S = 310\text{ly} = 95\text{ pc}$ its linear diameter $D = 2R_S = 4.1 \cdot 10^{11}\text{m}$ was obtained, which is 300 times larger than the diameter of our sun, which means that its volume is $2.7 \cdot 10^7$ times larger than that of the sun.

The technique was improved by Hanbury-Brown and Twiss [3], who used an intensity correlation method instead of the amplitude-interference, which is not sensitive against different air turbulences for the two interfering partial light rays incident onto the two mirrors M_1 and M_2. In this technique the number $(dN/dt)\Delta t$ of detected photons in the time interval Δt in the two channels is measured (Fig. 5.7). With a correlator the signal

$$S(\tau) = \int_0^{\Delta t} \frac{I_1(t) \cdot I_2(t+\tau)}{\langle I^2(t)\rangle} dt = \frac{1}{\langle \dot{N}^2\rangle} \int_0^{\Delta t} \frac{dN_1(t)}{dt} \cdot \frac{dN_2(t+\tau)}{dt} dt, \tag{5.7}$$

is measured, which depends on the delay time τ between the signals in the two channels.

The function $S(\tau)$ is a measure for the coherence of the detected star radiation and therefore, as in the Michelson experiment, a measure of the diameter of the star.

The most important method for the determination of star diameters is (before the most accurate data of Hipparcos and Gaia were available) the analysis of the light curves I(t) of photometric binaries, which will be discussed in the next section.

5.3 Binaries and the Determination of Masses and Radii of Stars

For the determination of masses of stars binaries turn out to be very useful sources of information. Such systems consist of two stars which orbit around their common center of mass S (Fig. 5.8). Binaries are by no means an exception among all stars. More than 50% of all stars belong to systems with two or more stars. Examples are the binaries *Alcor* and *Mizar* (Fig. 5.9) in the constellation *Ursus Major*.

The distance between the two components of binaries range from some 10^2 AU down to values only slightly

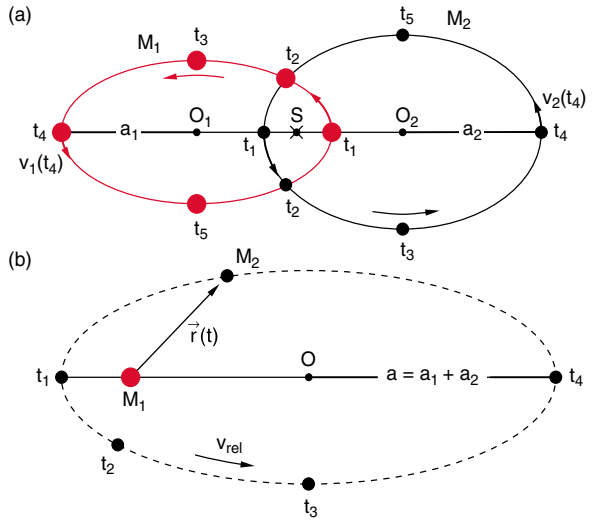


Fig. 5.8 Elliptical orbits of the two components of a binary system. **a** in the center of mass system **b** in relative coordinates with the origin in the heavier component M_1

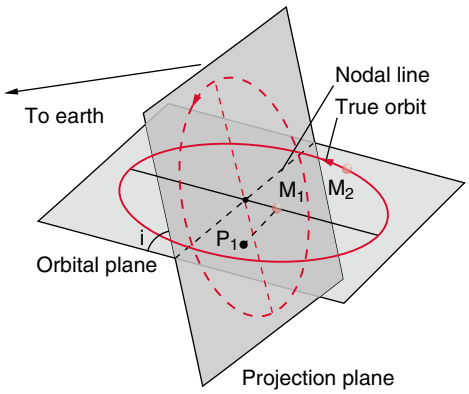


Fig. 5.9 True orbital plane of a binary system and projection of the plane onto a direction perpendicular to the line of sight to the earth. The projection P_1 of the point M is no longer located in the focal point of the projected ellipse

larger than the sum of the radii of the two stars, where they nearly touch each other. The revolution periods are related to their distance by the 3rd Kepler law. They range from a few hours for very small distances up to several thousands of years for large distances.

Both components move on elliptical orbits around the common center of mass S in Fig. 5.8a. In a coordinate system, where one component (generally the larger mass) sits in a focal point of the elliptical orbit of the other component (Fig. 5.8b), the distance between M_1 and M_2 is equal to the radius r(t) from the focal point to the orbiting second component M_2.

5.3.1 Visual Binaries

For visual binaries, where both components can be observed on their orbit around the common center of mass, the ratio of their masses

$$\frac{M_1}{M_2} = \frac{a_2}{a_1}. \tag{5.8}$$

equals the inverse ratio of the large semi-axis of their elliptical orbits. In the coordinate system with the origin at the center of the larger mass, is $a = a_1 + a_2$, which gives the relations

$$a_1 = \frac{M_2}{M_1 + M_2} a;\; a_2 = \frac{M_1}{M_1 + M_2} a \tag{5.9}$$

Often the masses of the stars are expressed in units of the sum of the two masses. From (5.9) we then obtain:

$$\begin{aligned} \frac{M_1}{M_\odot} &= \frac{M_1 + M_2}{M_\odot} \frac{a_2}{a}; \\ \frac{M_2}{M_\odot} &= \frac{M_1 + M_2}{M_\odot} \frac{a_1}{a}. \end{aligned} \tag{5.10}$$

The sum $M_1 + M_2$ of the two masses can be related to the revolution period T according to the third Kepler's law as

$$M_1 + M_2 = \frac{4\pi^2}{G} \frac{a^3}{T^2} \tag{5.11}$$

where G is the gravitational constant.

If a_1 and a_2 and T are known, the two masses can be obtained. The major semi-axis a appears from the earth under the angle $\alpha = a/r$ and has the parallax $\pi = 1\text{AU}/r$. The absolute value of a is then

$$a = \frac{1\text{AU}}{\pi} \cdot \alpha. \tag{5.12}$$

and the Eq. (5.10) becomes

$$\frac{M_1 + M_2}{M_\odot} = \left(\frac{\alpha}{\pi}\right)^3 \left(\frac{1 T_S^{(E)}}{T}\right)^2, \tag{5.13}$$

where $T_S^E = 1$ year is the siderial revolution period of the earth.

The relation (5.13) is only valid, if the orbital plane of the two components is either perpendicular or parallel to the line of sight binary- earth. Generally this plane is, however, inclined to the line of sight (Fig. 5.10). For the correct application of Eq. (5.13) one must determine the inclination angel i. This is possible because for the inclined situation

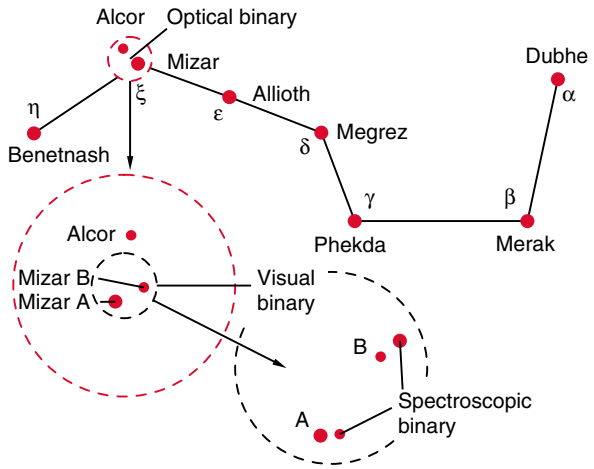


Fig. 5.10 Examples of optical binaries (Alcor and Mizar) in the constellation of the great bear, visual binaries (Mizar A and B) and spectroscopic binaries (each two components of Mizar A and Mizar B)

the larger mass in Fig. 5.8b does not appear exactly in the focal point of the elliptical orbit of the smaller component. The correct inclination angle is obtained, if the measured orbital data, projected onto the real orbital plane fulfil Kepler's laws.

5.3.2 Spectroscopic Binaries

The principle of the spectroscopic measurement of the orbital data of both components of binaries is based on the Doppler-shift of spectral lines, emitted by the two components. It is illustrated in Fig. 5.11. In the position A no Doppler-shift occurs, because the velocity of both components is perpendicular to the line of sight. In the position B the spectral lines emitted by the component 1 are shifted to the blue, those of component 2 to the red.

Example

The Doppler-shifts $\Delta\lambda(t)$ at $\lambda = 589\,\text{nm}$ of the two components A and B follow approximately a sine-function $\Delta\lambda(t) = \Delta\lambda_0 \cdot \sin\omega t$ (Fig. 5.12a) with a period $T = 2\pi/\omega$, which we have assumed for our example as $T = 11\text{d}$. The measured value of $\Delta\lambda_0$ are $\Delta\lambda_0(A) = 1.6 \cdot 10^{-4}\lambda_0$ and $\Delta\lambda_0(B) = 3.2 \cdot 10^{-4}\lambda_0$. Since $\Delta\lambda_0 = \lambda_0 \cdot v/c$ we can deduce from the measured values $\Delta\lambda$ the velocities of the components, moving on a circular orbit. They are obtained as $v_A = 50\,\text{km/s}$ and $v_B = 100\,\text{km/s}$ (Fig. 5.12b). Their distances from the common center of mass S is.

$$r_A = v_A/\omega = v_A \cdot T/2\pi = 5 \cdot 10^4 \cdot 1.5 \cdot 10^5\text{m} = 0.75 \cdot 10^{10}\text{m} = 0.05\text{AU},$$
$$r_B = v_B/\omega = 0.1\text{AU}.$$

◄

5.3.3 Photometric Binaries

The photometric binaries (eclipsing binaries) were for a long time a marvel for Astronomers, because one could not explain the fluctuations of their brightness. Only in 1782 J. Goodrike could show for the example of Algol (ß-Persei) that the fluctuations were periodical and he made the assumption that they were caused by the occultation of one component of a binyary system by the other component. This hypothesis was confirmed 100 years later by H. C. Vogel, who measured the periodic Doppler shifts of the emission lines for both components. In Fig. 5.13 the light curve for the apparent brightness of an ecliptic binary is schematically illustrated, where we have assumed that A is the brighter component and B the fainter one.

From the progression of rise and decrease of the light curve the ratio of the radii R_A/R_B can be deduced. Knowing the velocities of A and B from the Doppler-shifts, and the revolution perod T the individual masses M_A and M_B can be calculated. From the intensity drops at the times t_2 and t_4 the ratio of the luminosities of A and B can be deduced (see Problem 5.3), With telescopes with adaptive optics and even more accurately with the astrometric satellite GAIA, the two components can be resolved if their mutual distance is larger than the sum of their radii R_A and R_B [4].

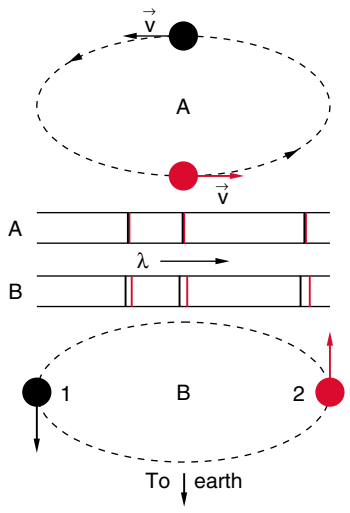


Fig. 5.11 Doppler-split of a spectral line for spectroscopic binaries

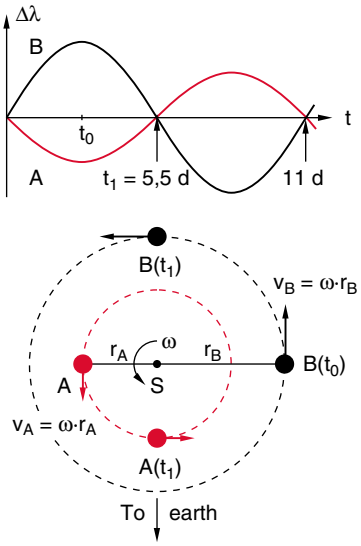


Fig. 5.12 **a** Analysis of the star orbits of a binary system from the Doppler shifts **b** Orbits of the components A and B

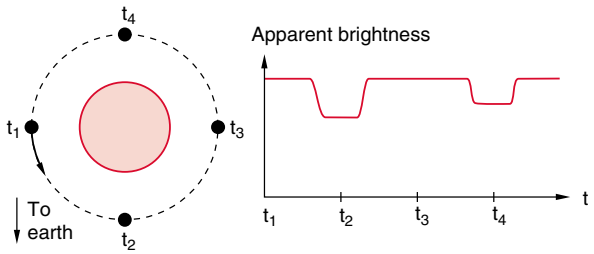


Fig. 5.13 Occultation binaries and their schematic light curve

In Fig. 5.9 the different types of Binaries are shown that can be observed in the constellation *Ursus Major.*

5.3.4 Spectral Types of Stars

Measuring the radiation of a star with spectral resolution, all stars show, as the sun. a continuous spectrum (see Sect. 4.5), superimposed by narrow dark absorption lines as well as bright emission lines. The spectral distribution of the continuum depends on the surface temperature of the star and the thickness of the layer emitting the continuum (photosphere). The absorption lines are generated by absorbing transitions between discrete states of atoms or molecules which have a lower temperature than the layer emitting the continuum, while the emission lines are due to transitions in atoms or molecules located at regions of higher temperatures, e.g. in the Chromosphere. The discrete line- spectrum gives information about the chemical composition of the absorbing or emitting layers of the star.

The first photographic recording of a star spectrum has been performed 1872 by *Henry Draper*. Later on the

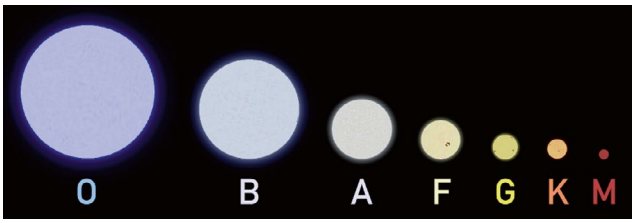


Fig. 5.14 Classification of Star types according to their surface temperature [Omnidoom 999—Own work, CC BY-SA 3.0, https://commons.wikimedia]

spectra of about 400.000 stars were recorded at the Harvard Observatory with a prism spectrograph and classified according to certain characteristic features. This **Harvard classification** [5] is based on the intensity ratio of relevant absorption lines in the star spectrum, as for instance the Balmer lines of the H-atom, the lines of the He-atom and of the iron atom, the Calcium-ion and several molecules, such as the CN-radical and the TiO molecule. These ratios depend critically on the temperature of the absorbing surface layers. There are several schemes for the classification of the different types of stars, which use either the luminosity or the spectral distribution of the emitted radiation.

The spectral types of the Harvard Classification are labelled by capital letters. In the sequence of decreasing temperatures, they are (Fig. 5.14)

$$\mathrm{O-B-A-F-G-K} \begin{matrix} \nearrow \mathrm{S} \\ \rightarrow \mathrm{M} \\ \searrow \mathrm{C} \end{matrix} \qquad (5.14)$$

Take home message: *Oh, be a fine Girl, kiss me* (smac!) (5.14).

The hottest stars are the O-stars, the M-stars the coldest (see Table 5.3) The additional stars types S and C were introduced after it was discovered, that some cooler stars showed absorption lines of molecular carbon C_2, of CH and CN-molecules, and of metal oxides such as ZrO and LaO in S-stars, which had not been found in M-stars with the same surface temperature.

Later on this classification was refined by the subdivision of each spectral type into 10 subgroups. For instance our sun belongs to the subgroup G2 [6].

The calibration of the Harvard-Sequence was fixed by the spectra of selected reference stars (Table 5.4) (Fig. 5.15).

5.3.5 Hertzsprung-Russel-Diagram

During the years 1910–1913 the Danish Astronomer *Ejnar Hertzsprung* (1873–1967), who was Professor at the Universities in Göttingen (Germany), Potsdam and Leiden

Table 5.3 Classification of Stellar types [Wikipedia]

Class	Effective temperature (K)	Vega-relative chromaticity	Chromaticity (D65)	Main-sequence mass (Solar masses)	Main-sequence radius (Solar radii)	Main-sequence luminosity (Bolometric)	Hydrogen lines	Fraction of all main-sequence stars (%)
O	≥30,000	Blue	Blue	≥16 $M_\odot$	≥6.6 $R_\odot$	≥30,000 $L_\odot$	Weak	~0.00003
B	10,000–30,000	Blue white	Deep blue white	2.1–16 $M_\odot$	1.8–6.6 $R_\odot$	25–30,000 $L_\odot$	Medium	0.13
A	7,500–10,000	White	Blue white	1.4–2.1 $M_\odot$	1.4–1.8 $R_\odot$	5–25 $L_\odot$	Strong	0.6
F	6,000–7,500	Yellow white	White	1.04–1.4 $M_\odot$	1.15–1.4 $R_\odot$	1.5–5 $L_\odot$	Medium	3
G	5,200–6,000	Yellow	Yellowish white	0.8–1.04 $M_\odot$	0.96–1.15 $R_\odot$	0.6–1.5 $L_\odot$	Weak	7.6
K	3,700–5,200	Light orange	Pale yellow orange	0.45–0.8 $M_\odot$	0.7–0.96 $R_\odot$	0.08–0.6 $L_\odot$	Very weak	12.1
M	2,400–3,700	Orange red	Light orange red	0.08–0.45 $M_\odot$	≤0.7 $R_\odot$	≤0.08 $L_\odot$	Very weak	76.45

Table 5.4 The luminosity classes (after Carroll and Ostlie 1996, p. 248)

Class	General Features	Examples
Ia-0	Extreme luminous super giants	ρ Cas RW Cep
Ia	Luminous super giants	α Ori β Ori
Ib	Less luminous super giants	α Car α Sco
II	Bright giants	θ Lyr α UMi
III	Normal giants	α Tau β Car
IV	Sub-giants	α CMi A η Boö
V	Main sequence (dwarfs)	α Cen A Sun
VI, sd	Sub-dwarfs	Prox Cen
D	White dwarfs	α CMa B 40 Eri B

(Netherland) and the American Astronomer *Henry Norris Russel* (1877–1957) in Princeton found, by comparing many different star spectra, a remarkable systematic relation between the absolute luminosity and the spectral classification of stars. Plotting the absolute luminosities of measured stars against their spectral class one obtains the diagram of Fig. 5.16 which is called the **Hertzsprung-Russel** Diagram HDR (sometimes also called *Color-Brightnes Diagram*).

From this diagram it is apparent that the stars can be divided in different grupes which represent (as will be shown below) different development stages of stars. The major fraction of all stars is located in the proximity of a curve, called the **main sequence**. It turns out that these stars are in their stable stage which belongs to the by far longest stage in their lifetime. Above the main sequence there are three ranges belonging to **Giant stars** ($R_S \gg R_\odot$). On the left side below the main sequence are the **White Dwarfs** (Fig. 5.17)

According to the MKK-classification (after William Morgan, Phillip Keenan, Edith Kellman (Yerkes Observatory) one distinguishes between different luminosity classes (Table 5.4).

In the range above the main sequence there are several variable stars with periodically or randomly changing luminosities which represent unstable transition stages of stars.

For the stable stars on the main sequence Arthur Stanley Eddigton (1882–1944) found a relation between the mass M of a star and its luminosity L (**mass-luminosity relation**)

$$L \propto M^\alpha. \tag{5.15}$$

For larger Masses ($M > 3\,M_\odot$) the exponent a is a ≈3, for small masses ($M < 0.5\,M_\odot$) is a ≈4. According to this relation a star with $M = 10\,M_\odot$ has a 1000 times larger luminosity than the sun.

The white Dwarfs and the red Giants deviate from this relation. The smallest masses observed so far are about $0.05\,M_\odot$. They are located in the Hertzsprung-Russel diagram in the lower right part of the main sequence. The largest masses, found up to now, in the range $10\,M_\odot$–$50\,M_\odot$ are in the upper right part of the HRD. One exception is the variable star η-Carina in the southern part of the Milky Way with a mass of $M = 100\,M_\odot$. The masses of the White Dwarfs are all smaller than $1.5\,M_\odot$.

In Fig. 5.18 the mass-luminosity relation is plotted in a logarithmic form.

Remark With the much more accurate data of the satellites *Hiparcos* and *Gaia* the masses and absolute luminosities of more than 100 million stars are now known much better.

Of particular interest is the observation of stars in a cluster, because all stars of the cluster have been formed at the same time. This gives the opportunity to study the relation between the mass of a star and its development stage.

We will now discuss the importance of these diagrams for our understanding of the development of stars.

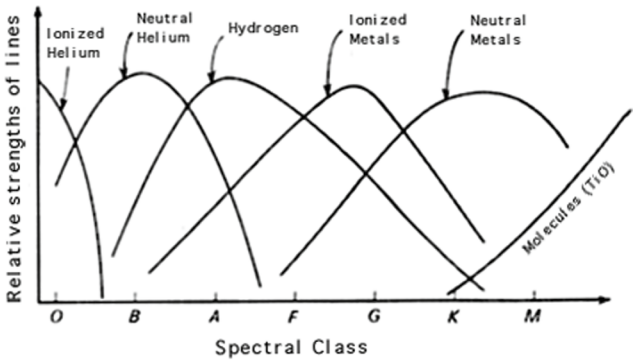


Fig. 5.15 Main contributing elements for the different classes of spectral types of stars

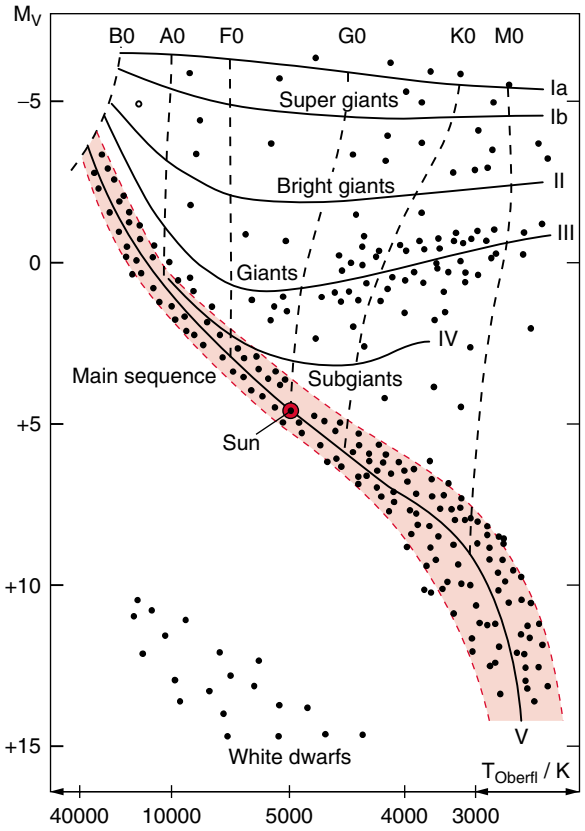


Fig. 5.16 Hertzsprung-Russel diagram (after Karttunen et al. Astronomy [8])

5.4 The Birth of Stars

According to our present knowledge stars ar created by the contraction of huge clouds of gas and dust. These clouds have been formed by interstellar matter. Their local density can have been enhanced by various mechanisms, for example by the explosions of stars and the shock waves created by such events (see Sect. 5.9). In our galaxy which contains about 10^{11} stars, new stars are mainly created in the spiral arms. The mean birth rate of stars is about 3–5 stars per year. The birth process lasts for 10^3–10^7 years, strongly dependent on the mass of the new star.

One example of an interstellar gas and dust cloud, where the birth of new stars can be observed, is the *Orion Nebula* (Fig. 5.19). Another example is the *horse head nebula* embedded in the vast and complex Orion nebula (Fig. 5.20). The gas clouds consist mainly of atomic and molecular hydrogen gas. Since the gas is cold, atoms can form molecules at collisions. Alhough such collisions are very seldom due to the low density, they have still a large chance to happen during the long lifetime of the cloud.

5.4.1 The Jeans Criterion

The contraction of a such a cloud with mass M and density ϱ can occur, if the attractive gravitational forces becomes larger than the repulsive gas pressure plus the centrifugal force, if the gas cloud rotates.

We will at first consider a nonrotating gas cloud with homogeneous mass density ϱ and a mean temperature $\overline{T}$. The mass density is so low, that we can regard the gas as ideal gas. This allows the application of the general gas equation

$$p \cdot V = \nu \cdot R \cdot T,$$

(R = gas constant, T = temperature, ν = number of moles). which gives for the pressure p of a gas of particles with mean atomic mass number $\overline{\mu}$ and atomic hydrogen mass m_H

$$p_{gas} = \frac{\varrho}{\overline{\mu} \cdot m_H} \cdot kT, \tag{5.16}$$

Example

For a cloud with hydrogen molecules H_2 is $\mu = 2$, for a cloud with 70% H-atoms, 10% H_2-molecules and 20% He-atoms is.

$$\overline{\mu} = 0.7 \cdot 1 + 0.1 \cdot 2 + 0.2 \cdot 4 = 1.7. \blacktriangleleft$$

For a spherical gas cloud with mass M and radius R the gravitational gas pressure is (see Problem 5.5)

$$p_{grav} = -\frac{3}{8}\frac{G \cdot M^2}{\pi R^4}. \tag{5.17}$$

A cloud collapse starts if $p_{gas} + p_{grav} < 0$. This gives the **Jeans Criterion** for the minimum mass of a collapsing gas cloud

$$M \geq \frac{2kT}{G \cdot \overline{\mu} \cdot m_H} \cdot R. \tag{5.18}$$

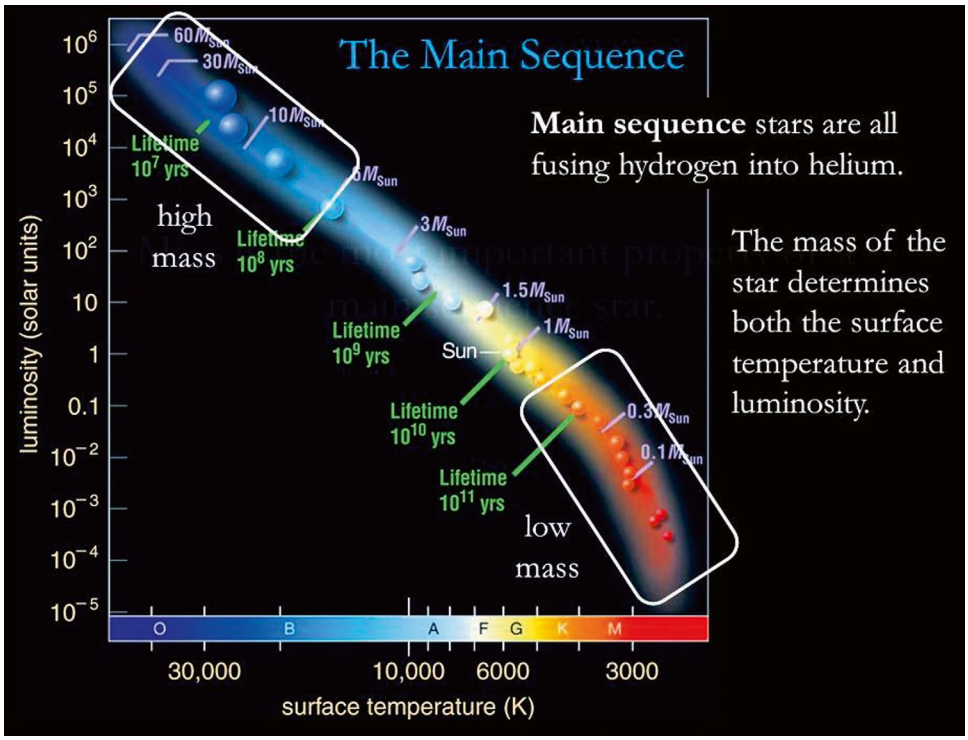

The Main Sequence
Main sequence stars are all fusing hydrogen into helium.
The mass of the star determines both the surface temperature and luminosity.
luminosity (solar units)
surface temperature (K)
high mass
low mass
Lifetime 10^7 yrs
Lifetime 10^8 yrs
Lifetime 10^9 yrs
Lifetime 10^10 yrs
Lifetime 10^11 yrs
Sun
O B A F G K M
30,000 10,000 6000 3000

Fig. 5.17 The main sequence in the Hertzsprung-Russel Diagram (Wikipedia)

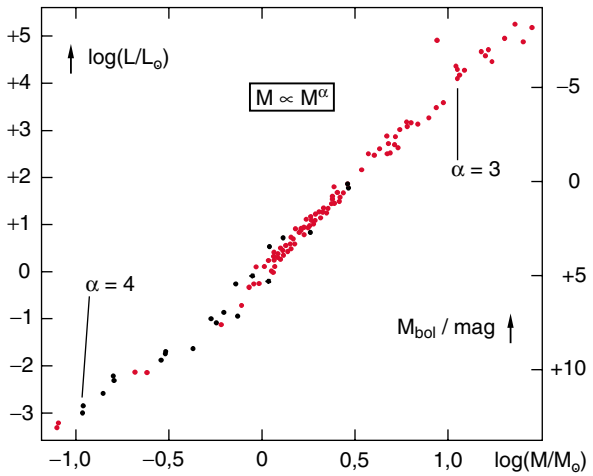


Fig. 5.18 Mass-Luminosity diagram. Red points: spectroscopic binaries. Black points: visual binaries (from Unsöld-Baschek: Der Neue Kosmos Springer Heidelberg 1991)

Example

Inserting typical numerical values for a diffuse interstellar cloud of atomic hydrogen gas with $T = 100\text{K}$; $R = 20\text{pc} = 6 \cdot 10^{16}\text{m}$; $\overline{\mu} = 1$ one obtains $M > 2 \cdot 10^4\, M_\odot$. ◄

Many clouds have only masses of 50–300 $M_\odot$. If the other values are the same as in the example, they cannot collapse, because their repulsive thermal energy is larger than the attractive gravitational energy.

The probability for a collapse increases with decreasing temperature of the cloud (Table 5.5).

Example

A cold molcular cloud of H_2 gas with $\overline{\mu} = 2$; $T = 20\text{K}$; $R = 20\text{pc} \rightarrow M > 2 \cdot 10^2\, M_\odot$. ◄

With $M = \frac{4}{3}\pi\overline{\varrho}R^3$ (5.18) can be written as relation between the mean density $\overline{\rho}$ of the cloud and its radius R

$$\overline{\varrho} = \frac{3kT}{2\pi \cdot G \cdot \overline{\mu} \cdot m_\text{H}} \cdot \frac{1}{R^2}$$
$$\Rightarrow R > \left(\frac{3kT}{2\pi \cdot G \cdot \varrho \cdot \overline{\mu} \cdot m_\text{H}}\right)^{1/2}. \qquad (5.19)$$

For $R = 20$ pc; $T = 20$ K one obtains a minimum density of 2 H_2-molecules per cm^3. The radius R of such a molecular cloud is about 10 times larger than the mean distance between stars in our Galaxy. Replacing R by $R = (3M/4\pi\varrho)^{1/3}$ Eq. (5.19) transfers into

$$M \geq \sqrt{\frac{6}{\pi}}\left(\frac{kT}{G \cdot \overline{\mu} \cdot m_\text{H}}\right)^{3/2} \cdot \frac{1}{\sqrt{\varrho}}. \qquad (5.20)$$

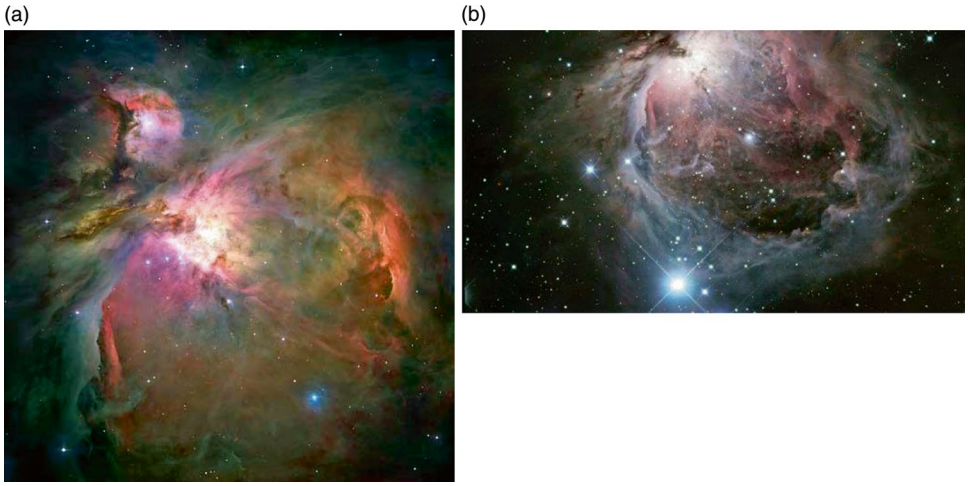


Fig. 5.19 Orion nebula. **a** photo taken in the visible spectrum by the Hubble Telescope (NASA). **b** photo taken in the infrared region with the infrared telescope on the Paranal in Chile (ESO)

Fig. 5.20 The horse head nebula in Orion (NASA)

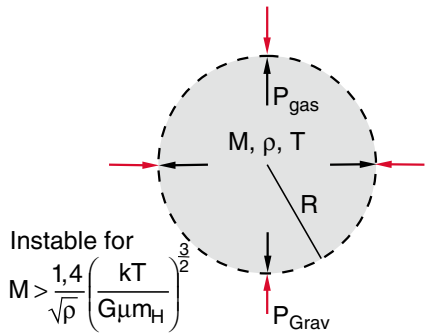


Fig. 5.21 Contraction of a gravitationally unstable cold gas cloud

Table. 5.5 Minimum mass (in units of $M_\odot$) of a gravitative unstable gas cloud for different concentrations of H-atoms

T	1 H-atom per cm^3	10^2 H-atoms per cm^3	10^4 H-atoms per cm^3
5 K	270	27	2.7
10 K	750	75	7.5
100 K	25.000	2500	250

The minimum mass M_{min} of a cloud undergoing a gravitational collaps is independent of its radius R, but depends only on its density ϱ and its temperature T (Fig. 5.21).

In Table 5.5 the minimum masses are listed for different values of density ϱ and temperature T.

Note: The Jeans criterion can be also derived from the Virial theorem. For an attractice force proportional to r^{-2} a gas is at stable thermodynamic equilibrium if twice of the mean kinetic energy $2\overline{E}_{kin} = 2 \cdot \frac{3}{2}kT$ is equal to the absolute value of the mean potential energy (Note that $E_{pot} < 0$)

$$2\overline{E}_{kin} = -E_{pot} = |E_{pot}|. \tag{5.21}$$

The cloud becomes unstable if $2\overline{E}_{kin} < |E_{pot}|$. This (5.21) gives the same order of magnitude for the instability as the Jeans Criterion.

These estimations have shown, that collapsing gas clouds have a minimum mass of some hundred solar masses, depending on the mean density (Table 5.5). Since the stars observed so far have masses $M < 50\,M_\odot$ this implies that at the contraction of gas clouds not single stars emerge but star clusters, i.e. many stars which are born simultaneously. The other possibility is, that not the whole gas cloud condenses into stars but only part of it.

It turns out that both possibilities happen. At the formation of star clusters part of the contracting gas cloud remains as interstellar gas.

The formation of star clusters can be explained, if one assumes that during the collapse of the cloud local

variations of the density emerge. Such variations can occur, for instance, by shock waves, which are produced when stars eject their outer gas shell during the end of their lifetime. In particular shockwaves which are generated during supernovae explosions of massive stars can result in local compressions of the density in gas clouds.

In the spiral arms of galaxies density waves are produced by the differential rotation of the galaxy (see Sect. 6.4). Also intrinsic phenomena can cause local density variations such as local magnetic fields, turbulences during the collapse and friction.

Such locally limited part of the cloud with higher density ϱ can become gravitationally unstable, since the minimum radius R of such a part decreases with ϱ as $R \sim \varrho^{-1/2}$.

Such clouds with locally differing densities form individual condensation areas, i.e. the collapsing gas cloud fragmentates into localized parts which become the predecessors of single stars.

We will now estimate the minimum collaps time from the beginning of the contraction of the gas cloud until the formation of a proto-star:

A particle with mass m and zero initial velocity at the distance r from the center of the spherical gas cloud experiences the attractive gravitational force

$$F = -G \cdot M(r) \cdot m/r^2 \tag{5.22}$$

and the acceleration (which depends on r)

$$\ddot{r}\,(r) = G \cdot \frac{4}{3}\pi\wp r^3/r^2 = G \cdot \frac{4}{3}\pi\wp r \tag{5.23}$$

where $M(r)$ is the mass of the cloud inside the radius r. Neglecting collisions the motion of the particle is a sinusoidal oscillation with the frequency $\omega = \left(\frac{4}{3}\pi\varrho G\right)^{1/2}$ (see Vol. 1, Problem 2.26).

The falling time is ¼ of the oscillation period $T = (2\pi\omega)^{-1}$.

$$\tau_{\text{fall}} = \frac{2\pi}{4\omega} = \frac{1}{4}\sqrt{\frac{3\pi}{\varrho \cdot G}} = \frac{1}{2}\pi\sqrt{\frac{\omega^3}{G \cdot M(r)}} \tag{5.24}$$

Example

$n = 10^2$ H -atoms/cm^3 $\Rightarrow$ $\varrho = 1.67 \cdot 10^{-19}$ kg/m^3 ,
$G = 6.67 \cdot 10^{-11}\text{m}^3\text{kg}^{-1}\text{s}^{-2} \Rightarrow \tau_{\text{fall}} = 2.3 \cdot 10^{14}\text{s}$
$= 7.3 \cdot 10^6$a. ◀

This illustrates that the collapse-times and therefore the formation times of stars are very long, compared with the lifetime of human beings. **Stars do not emerge suddenly.**

As Eq. 5.24 shows, the collapse time depends on the mass M(r) of the contracting cloud. In Table 5.6 the collapse times for different masses of the contracting gas cloud are listed.

Table. 5.6 Collapse-times for different masses of the protostar

$M/M_\odot$	$\tau/10^6$ years
0.5	150
3	1.050
9	0.15
15	0.06

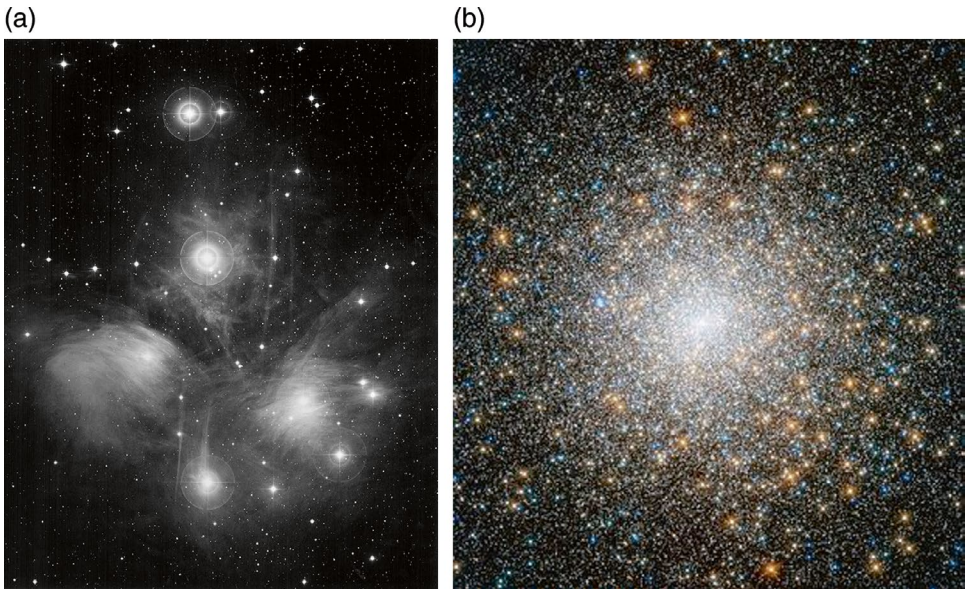


Fig. 5.22 **a** The Plejades in the constellation Taurus. The rest of the interstellar matter from which the stars have been formed can still be seen as light scattering nebula around the stars. **b** Star cluster Messier 68 = NGC 4590 is 36.000 Ly away from the earth (Wikipedia)

5.4.2 The Formation of Proto-Stars

We will now look in more detail into the collaps of gas-and dust clouds. During the collaps the density ϱ and therefore also the pressure p increase continuously, while the potential energy decreases (it becomes more negative). As long as the density is sufficiently small, the released potential energy ΔE_{pot} can be radiated away since the absorption of this radiation is small because of the small density ϱ. The temperature of the cloud does therefore not increase substantially. This means that in this early stage of the contraction the increasing density occurs isothermal. The Jeans Mass $M_{\text{J}} \propto \varrho^{-1/2}$ decreases and therefore subsets of the total cloud can contract towards their own mass center, if local density fluctuations become large enough.

Under these conditions stars therefore are not formed as single individuals but as star clusters.

All stars of the cluster have the same age. They form an interesting research subject, because one can study the mass-dependent development of stars, that have been born at the same time. One example of an open star clusters are the Plejades (Fig. 5.22a) which contain only a few stars. The bright remains of the gas cloud can be still seen as luminous gas around the new stars.

Another example is the globular cluster Messier 68 (Fig. 5.22b) which contains several thousand stars. The Messier Catalogue, which has been compiled by Charles Messier (1730–1817) was for a long time used as the standard galaxy and star catalogue, but has now been replaced by the NGC (new great catalogue).

When the contracting cloud becomes with increasing density optically opaque the radiation cannot escape the cloud but contributes to the heating of the cloud. If the thermalization time (i.e. the mean time between atomic collisions) is short compared to the collapse time the contraction and heating of the cloud is adiabatic: The temperature increases. The pressure $p \propto \varrho kT$ increases more strongly than the density ϱ. The temperature rise is larger in the central part of the cloud than at the edges, because the radiation has a lower chance to escape from the center than from the edges. Therefore in the central part a region of high pressure and temperature is formed. When the gas pressure becomes equal to the gravitational pressure, the collapse of the central part ends and it becomes stabilized.

The atoms from the outer isothermal part, however, continue to fall onto the central part and heat it further up. Due to the increasing temperature the pressure also increases and the Jeans-mass decreases. Depending on the size and the temperature of the contracting part of the cloud, central masses between 0.1 $M_{\odot}$ and 50 $M_{\odot}$ are formed. This agrees with the observed masses of stars.

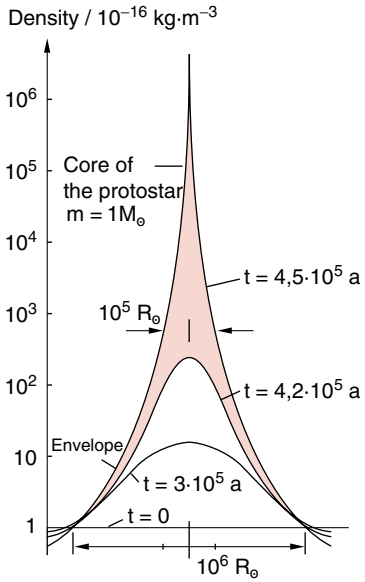


Fig. 5.23 Radial density progression of a contracting gas cloud with initial density $\rho = 10^{-16}$ kg/m, temperature T = 10 K and mass $M = M_{\odot}$ for different contraction times (from H. Scheffler, H. Elsässer: Physik der Sterne und der Sonne Wissenschafts-Verlag Mannheim 1990)

The spherical gas contraction achieved at the end of the collapse forms a central part with sufficient high density to prevent the escape of any radiation and which has therefore a very high temperature. This central part is called **protostar** (Fig. 5.23).

The fast adiabatic pressure rise (no energy release!), following the relation

$$p \cdot V^{\kappa} = \text{const.} \tag{5.25}$$

retards the collapse. This retardation starts at a temperature of about $T = 100\text{K}$ and a density $\varrho = 10^{-7}\text{kg/m}^3$. During this slower contraction the temperature in the central part rises to several thousand degrees.

At such high temperatures the molecular hydrogen H_2 starts to dissociate. The necessary dissociation energy decreases the kinetic energy of the particles. This results in a retardation of the temperature and pressure increase (Fig. 5.24), while the gravitational pressure continues to rise unrestraint due to the increasing density.

The protostar contracts further until all molecules are dissociated. Now the temperature increases again faster. The central part, however, is not visible because the emitted radiation is absorbed in the optically thick outer layers. This absorption heats the envelope of the protostar up to about $T \approx 700\text{K}$. It radiates in the infrared region ($\lambda(I_{\max}) > \approx 4\,\mu\text{m}$).

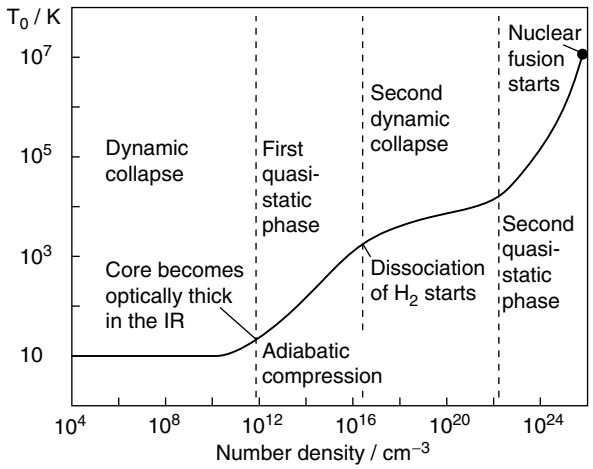


Fig. 5.24 Schematic progression of the central temperature as a function of the particle density in the center during the collapse of a molecular cloud fragment until the formation of a star on the main sequence (from H. Scheffler, H. Elsässer: Physik der Sterne und der Sonne Wissenschafts-Verlag Mannheim 1990)

The protostar appears as object in the infrared

Due to the gravitational attraction more and more particles from the envelope fall onto the central region. This makes the envelope optically thinner and the radiation from the central part can now penetrate through the envelope which makes the core visible. It appears as infrared radiation source. Further contraction lets the core temperature rise further to a value, where most of the H-atoms are ionized. The necessary ionization energy is again taken from the kinetic energy of the particles. The temperature rise and the contraction velocity is slowed down. When the central temperature has reached $T = 10^5$K the gas is totally ionized. The core reaches hydrostatic equilibrium when the thermal pressure of the ionized gas becomes equal to the gravitational pressure. The radius of the protostar has decreased during this period from 100AU $\left(10^4 R_\odot\right)$ to about $0.2\text{AU}(40R_\odot)$.

Since the shell of the contracted cloud has become so thin, it can no longer absorb the radiation emitted by the core. It is therefore no longer heated by the absorption of the radiation. Its kinetic energy stays below the potential energy and it falls down onto the core. This increases the mass of the core and with it the gravitational pressure. The core continues to contract slowly and its temperature rises furthermore [7].

5.4.3 The Influence of Rotation on Contracting Gas Clouds

Up to now we did not take into account, that most of the gaseous clouds rotate. If, for instance, a cloud is formed in one of the spiral arms of our galaxy the particles in the cloud participate in the differential rotation of the galaxy (Fig. 5.25a). The contracting gas cloud has therefore an angular momentum relative to its center. Also during the fragmentation caused by turbulent vortices the different fragments carry angular momentum.

Since the angular momentum of a cloud, which is no longer in contact with its surroundings, is preserved, the angular velocity of the rotating cloud increases with increasing contraction. The collapse perpendicular to the rotation axis is retarded by the centrifugal force. Therefore an oblate ellipsoid develops. If a spherical protostar should emerge, angular momentum has to be transferred from the center to the edges of the cloud. This can be realized in different ways.

It turns out, that in the interstellar space a weak magnetic field exists ($B \approx 10^{-10}$T)) Since the collapsing gas cloud contains also ionized particles, the magnetic field lines are carried with the charged particles and are deformed in the rotating cloud (Fig. 5.25b). This slows down the rotation.

With increasing density the recombination rate of electrons and ions increases because collisions become more frequent. This decreases the ion density in the central part. In the outer parts of the cloud the charged particles are accelerated by the magnetic field which is carried along with the rotating particles. By this means angular momentum is transferred from the inner to the outer part of the cloud using a combination of magnetic field effects and collisions.

Such a process must have played an important role during the formation of our solar system because the sun carries 99% of the total mass of the system but less than 1% of the total angular momentum of all planets.

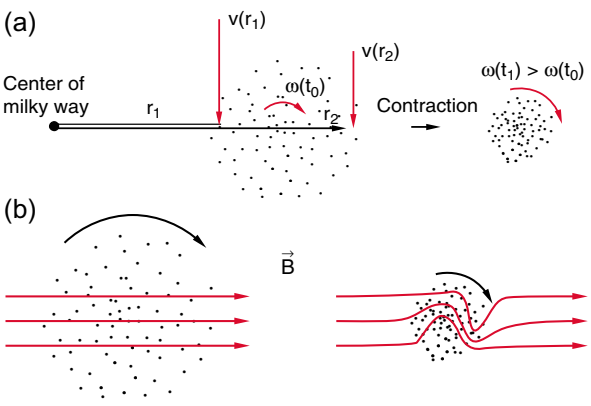


Fig. 5.25 **a** Illustration of the collapse of a rotating gas cloud. **b** Deformation of the magnetic field lines of the interstellar magnetic field by the contraction of a rotating gas cloud containing charged particles

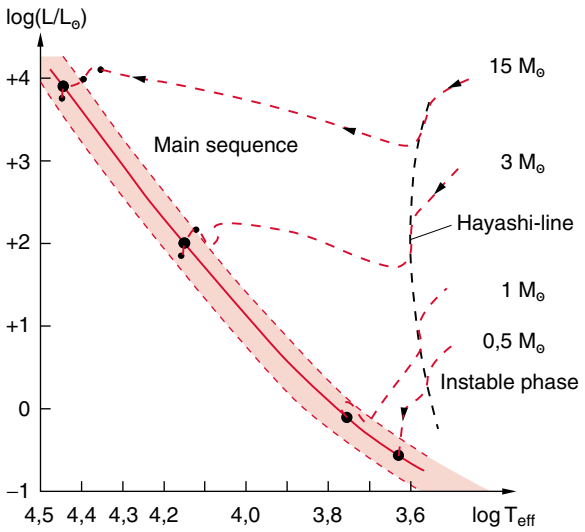


Fig. 5.26 Paths in the HRD of stars with different masses in their early state until they reach the main sequence

Fig. 5.27 Bright protoplanetary disc in the middle of the picture around a T-Tauri star (C. R. O' Dell/Rice University; NASA, Wikimedia)

5.4.4 The Progression of a Star into the Main Sequence of the Hertzsprung-Russel Diagram

When the protostar has reached its equilibrium stage, its temperature is still too low to ignite nuclear fusion processes. The energy transport from the central core to the outer layers is mainly realized by convection, which is very effective. The temperature gradient is smaller than that in our sun (see Sect. 4.5), This implies that the surface temperature is relatively high. Its place in the H.-R.-diagram is on the lower right of the main sequence and depends on its mass (Fig. 5.26). The Japanese Astrophysicist C. Hayashi calculated the condition for the equilibrium of convective stars, which gives a relation between luminosity L and surface temperature T_{eff} and depends on the mass of the protostar. In the HRD this condition represents a nearly vertical line (Hayashi-Line, Fig. 5.26). Stars located to the right of this line have not yet reached their equilibrium state.

During the development of a protostar to a stable star the protostar moves in the HRD from the right side towards the Hayashi line where it reaches its equilibrium state. The larger its mass is, the higher it meets the Hayashi line. During the following small contraction its surface area shrinks, while its surface temperature rises only slowly. Therefore its luminosity

$$L = 4\pi R^2 \sigma \cdot T^4 \tag{5.26}$$

decreases. It moves on the Hayashi line downwards.

The contraction increases the central temperature. When it reaches a critical value, nuclear fusion processes start and radiation transport now becomes the main energy transport mechanism. The star leaves the Hayashi line and moves towards the main sequence. During this intermediate state the energy production in the central part changes from contraction (gravitational energy) to the start of nuclear fusion and the star may pass unstable states (**T-Tauri phase**), where its temperature and radius undergo irregular fluctuations. This can be observed by measuring the Doppler-shift of spectral lines. Such measurements hint to radiation emitting gas which falls down onto the core with high velocity and also to turbulences in the motion of this gas. These observations have been made on T-Tauri stars embedded in diffuse nebulae where the star formation takes place (Fig. 5.27).

T-Tauri stars are a class of variable stars which are not older than 10 Million years. Their name comes from one representative in the tauri star constellation.

After these birth pangs the stable phase of the star begins. This is by far the longest evolution stage in the star life. The stable star stays on the main sequence in the HRD for a long time, depending on its mass (see Sect. 4.x). Therefore the main sequence contains the largest fraction of all stars observed today.

5.5 The Stable Stage on the Main Sequence

The lifetime of a star and its development stages are essentially determined by two characteristic features: **Its mass** and **its chemical composition**.

Protostars with masses $M < 0.08\,M_\odot$ do not have enough gravitational energy to reach sufficiently high central temperatures for the ignition of nuclear fusion processes. Therefore they do not contain enough energy supply for emitting radiation over a longer time. They do not reach

the main sequence in the HRD. They contract until the gas pressure can compensate the gravitational pressure. This compression results in a slight temperature increase which enhances the luminosity. Since this energy supply provided by the gravitational energy is soon exhausted, they cool slowly down and become cold celestial objects (**brown dwarfs**), which no longer emit radiation.

Protostars with masses $M > 50\,M_\odot$ have on the other side so much gravitational energy that their compression results in an extremely high temperature. Therefore besides the gas pressure the radiation pressure $\propto T^4$ (!) plays an essential role in compensating the gravitational pressure.

This radiation pressure drives the particles of the protostar towards the edge and the star loses matte. Furthermore the high temperature enhances the nuclear fusion processes which therefore proceed very fast and the lifetime of such massive stars is very short.

The main sequence of the HRD therefore contains stars with masses M between $0.08\,M_\odot < M < 50\,M_\odot$.

5.5.1 The Dependence of Luminosity and Lifetime of Stars on their Mass

The investigation of many stars on the main sequence has brought the empirical relation between the mass M of a star and its luminosity L (Fig. 5.28). *Arthur Stanley Eddington* summarized these experimental findings in a general **mass-luminosity-relation**

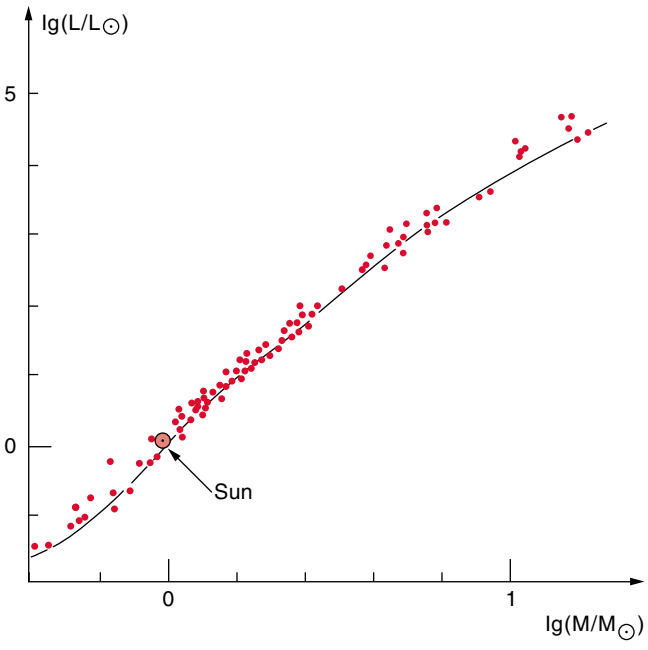


Fig. 5.28 Mass-Luminosity relation for stars on the main sequence

$$L \sim M^\alpha \tag{5.27}$$

where the exponent α has a numerical value between 3 and 4. As the double-logarithmic diagram of Fig. 5.28 shows, the slope of the curve L(M) is α = 3.7 for stars with masses M in the interval $M_\odot < M < 8\,M_\odot$ and decreases to α = 3 for larger masses. Also for stars with masses $M < 0.5\,M_\odot$ the slope becomes α = 3.

This relation allows the estimation of the average time span of a star on the main sequence.

It turns out that during this time the energy production in the central part of a star is essentially supplied by the fusion of Hydrogen to Helium. Detailed star models show, that during the stable burning phase inside a star, which contains mainly hydrogen, only about 10% of its hydrogen content undergoes fusion to Helium.

After this stable stage other fusion processes start which result in unstable conditions and the star leaves the main sequence.

The fusion energy produced per H-atom in the p-p-process (see Sect. 4.xx) is $E_\mathrm{H} = 5\,\mathrm{MeV} = 8 \cdot 10^{-13}\mathrm{J}$.

The dwell time on the main sequence is then

$$\tau_\mathrm{ms} = \frac{\eta \cdot (M/m_\mathrm{H}) \cdot E_\mathrm{H}}{L}, \tag{5.28}$$

where L is the luminosity, i.e. the emitted radiation power.

For stars with masses $M \approx M_\odot$ where the energy transport from the central part to the surface is performed mainly by radiative transfer and only partly by convection, (see Sect. 4.5) is $L \propto M^3$. Inserting this into (5.28) we get for the dwelling time on the main sequence

$$\tau_\mathrm{ms} \propto \frac{1}{M^2}. \tag{5.29}$$

The lifetime of stars with large masses is much shorter than that of stars with small masses.

Generally, the mass M and the luminosity L are expressed in units of the corresponding values of our sun. This modifies (5.28) into

$$\tau_\mathrm{ms} = \frac{\eta \cdot E_\mathrm{H}}{m_\mathrm{H}} \cdot \left(\frac{M/M_\odot}{L/L_\odot}\right) \cdot \frac{M_\odot}{L_\odot}, \tag{5.30}$$

Using the relation $L/L_\odot = (M/M_\odot)^3$ for stars with medium masses $M \approx M_\odot$ we obtain

$$\tau_\mathrm{ms} = \frac{\eta \cdot E_\mathrm{H}}{m_\mathrm{H}} \cdot \frac{M_\odot}{L_\odot}\left(\frac{M_\odot}{M}\right)^2. \tag{5.31}$$

Inserting the numerical values $E_\mathrm{H} = 5$ MeV; $m_\mathrm{H} = 1.67 \cdot 10^{-27}$ kg; $M_\odot$; $\eta = 0.1$ we get.

$= 2 \cdot 10^{30}$ kg; $L_\odot = 3.85 \cdot 10^{26}$ W

$$\tau_{ms} = 6 \cdot 10^9 (M_\odot/M)^2 \text{years}. \quad (5.32)$$

A star with $M = M_\odot$ stays, according to this model, for $6 \cdot 10^9$ years on the main sequence in the HRD, which means that its stable burning stage lasts for 6 Billion years.

One should compare this with the relatively short period of $5 \cdot 10^6$ years the star needs for the contraction phase and the protostar phase which is less than 1/1000 of its stable burning phase.

During the stable hydrogen burning phase unconsumed hydrogen from outer layers can be brought by convection into the central part where the fusion takes place, thus increasing the supply of fusion material and prolonging the stable burning phase. More accurate star models give a longer stable burning phase of the sun as $\tau_{ms} = 10^{10}$ years. Since the sun radiates already for 5 billion years there is still a time of 5 billion years left before the sun becomes unstable.

On the other hand a more massive star with $M = 20\ M_\odot$ stays only for 15 million years in its stable phase, which is only slightly longer than its birth phase.

5.5.2 Energy Production in Stars on the Main Sequence

The detailled features of the energy production and the energy transport to the surface as well as the time development of the star depend crucially on its mass. For stars on the lower part of the mains sequence ($M < 1.5\,M_\odot, T_{center} < 18 \cdot 10^6$ K) the energy is, like in our sun, mainly provided by the p-p reaction (see Sect. 4.5.4). With respect to the energy transport and the further development, stars with $M < 0.25\ M_\odot$ behave differently than stars with $M > 0.25\ M_\odot$.

5.5.2.1 $0.08M_\odot < M < 0.25M_\odot$

In stars with very small masses only a small central part of the star reaches sufficiently high temperatures to ignite fusion processes. The radial temperature gradient is then very large. If it becomes larger than the adiabatic temperature gradient

$$\left(\frac{dT}{dr}\right)_{adiabtic} = \left(1 - \frac{1}{\kappa}\right)\left(\frac{T}{p}\right)\frac{dp}{dr} \text{with } \kappa = \frac{c_p}{c_v} \quad (5.33)$$

(see Sect. 4.5.7) the radial gas layering becomes unstable and hot gas rises from the hot inner part to cooler outer layers. In this case convection represents the major part of energy transport (Fig. 4.107 and 5.29a). The convection intermingles the different layers of the star and the total supply of hydrogen is available for fusion. These stars move during their lifetime slowly on the main sequence up to the left. When the total supply of hydrogen is burnt to helium, pressure and temperature in the central part decline and the star collapses. During this collapse the temperature increase is not high enough to ignite the CNO- fusion cycle (see Sect. 4.5.8). The star slowly cools down and becomes a **White Dwarf** (see Sect. 5.x). Since there is no longer energy production in their inner part their luminosity decreases continuously and they finally become cold objects with zero luminosity.

5.5.2.2 $0.25\ M_\odot < M < 1.5M_\odot$

For stars with masses between 0.25 $M_\odot$ and 1.5 $M_\odot$ the situation is more complex. Here ignition temperatures for the p-p-reaction are reached for a larger central region ($r < 0.3R$) . The energy production is therefore distributed over a larger volume, which implies that no large temperature gradients occur in the central part and convection does not play a significant role (Fig. 5.29b). Energy transport in the central part occurs mainly by radiative transfer (see Sect. 4.y).

In the outer layers the temperature gradient dT/dr becomes larger because the absorption coefficient for radiation from the inner part increases due to the lower temperature. This favors convection.

The core of these stars is radiative, the outer layers convective. Therefore there is nearly no intermixing between core and outer regions. In the central part the hydrogen supply decreases, the helium share increases. In the outer layers the amount of hydrogen stays essentially constant.

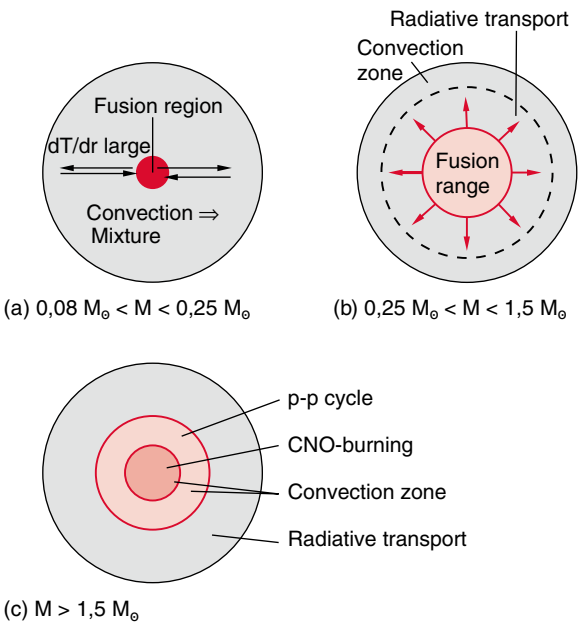


Fig. 5.29 Schematic illustration of energy production and energy transport in stars with different masses on the main sequence: **a** $M < 0.25\ M_\odot$ **b** $0.25\ M_\odot < M < 1.5\ M_\odot$ **c** $M > 1.5\ M_\odot$

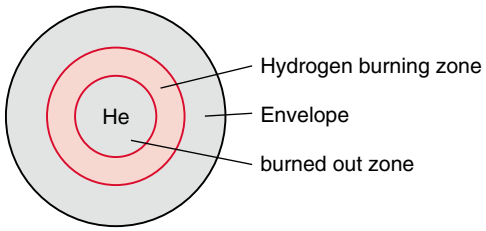


Fig. 5.30 Non-burning He-core and hydrogen-burning shell in the later state of the main-sequence development of stars with medium masses

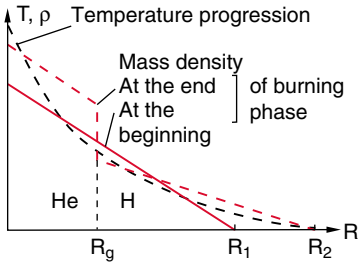


Fig. 5.31 Illustration of the increase of the radius R of massive stars at the end of the hydrogen burning

Due to the energy transport the surface, the temperature of the outer layers of the star increases and the star moves slowly on the main sequence to the upper left (Fig. 5.26).

When the hydrogen in the core is burnt to Helium the nuclear p-p fusion goes out, temperature and pressure decrease and the inner part of the star contracts. This causes a compression of the hydrogen containing layers around the core, their temperature increases and fusion processes start in these layers. The star gains an energy productive hydrogen burning layer surrounding the core of the star (Fig. 5.30).

During this stable burning phase the number of particles is diminished, because 4 hydrogen nuclei fuse to 1 He nucleus, but the mean particle mass $\overline{\mu}$ is increased. At the border shell $R = R_g$ between helium enriched core and hydrogen-rich layers the boundary conditions

$$p_{in} = p_{out}; \quad T_{in} = T_{out}. \tag{5.34}$$

must be fulfilled for hydrostatic equilibrium. With the general gas equation

$$p_{out} = N_{out} k T_{out}; \quad p_{in} = N_{in} k T_{in} \tag{5.35}$$

it follows for the particle density: $N_{in} = N_{out}$.

Since the total number of particles in the core decreases due to the fusion, the volume of the core must shrink. This increases the density in the hydrogen-rich zone with $r < R_g$, because the temperature is now higher than before the ignition of the hydrogen burning.

The mass density $\varrho = N \cdot \overline{\mu} \cdot m_H$ makes a step at the boundary between core and outer shell (Fig. 5.31). Since the total mass

$$M = 4\pi \int_0^R \varrho(r) r^2 \mathrm{d}r \tag{5.36}$$

of the star remains constant, the volume of the star must increase, i.e. its radius becomes larger. This results in a slight decrease of the surface temperature T_{eff}. But since the surface increases, the luminosity

$$L = 4\pi R^2 \cdot \sigma \cdot T_{eff}^4 \tag{5.37}$$

can still increase, if the increase of the factor R^2 over-compensates the decrease of T^4. The star moves in the HRD in Fig. 5.26 upwards to the right, leaving the main sequence.

This unstable stage starts when about 12% of the total star mass has been converted into helium (Fig. 5.32).

5.5.2.3 $M > 1.5 M_\odot$

In stars with masses above 1.5 $M_\odot$ in the upper part of the HRD in Fig. 5.26 the central temperature is so high, ($T > 20 \cdot 10^6$K) that the CNO-fusion process (see Sect. 4. xx) contributes the major part to the energy production (Fig. 4.98). Because of the steep temperature dependence T(R) on the distance r from the center, the main part of the energy production is restricted to a small volume around the center $r = 0$. The energy density in the central part and the energy flux to the outer part are very large, which causes a high radial temperature gradient $\mathrm{d}T/\mathrm{d}r$ (Fig. 5.29c). Therefore convection becomes the major contribution to the energy transport. This causes an effective intermingling of the star material and provides the central part, where fusion takes place with new hydrogen from the outer layers and ensures that the hydrogen supply decreases uniformly in the larger central part of the star.

In the outer layers outside the core the radial temperature gradient becomes smaller with the effect that here radiation represent the major energy transport. In the transition region between core and envelope exists a zone where the hydrogen content decreases with increasing r and radiative as well as convection undertake the energy transport.

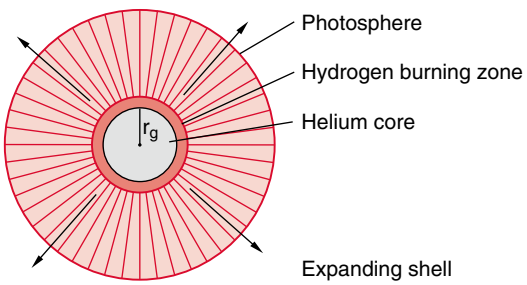


Fig. 5.32 State of a star after the consumption of hydrogen in its core

Such massive stars are convective in their central part while in the outer layers radiative energy transport takes over. This is inverse to the situation in stars with medium masses.

5.6 Star Development After Leaving the Main Sequence

The star development at the end of the stable burning stage, i.e. the point in the HRD where the star leaves the main sequence, depends critically on the mass of the star. The star mass also determines the time interval, the star spends on the main sequence, and the specific path in the HRD until the unstable stages of the star begin, (Fig. 5.33). Massive stars pass much faster through the development stages, indicated in Fig. 5.33 than stars with smaller masses.

If the hydrogen amount in the core has decreased to such a small fraction, that the energy production decreases distinctly, temperature and pressure in the core will decline and the core will contract. This contraction causes a temperature rise at the edge of the core which reaches sufficient high values to ignite hydrogen fusion processes in a shell around the core which still has enough hydrogen. This increases the temperature of the shell. Since the temperature of the core has decreased the radial temperature gradient becomes smaller. These processes proceed sufficiently slow to ensure always hydrostatic equilibrium. Under this assumption the Virial theorem is valid, which states

$$\Delta E_{\text{kin}} = -\frac{1}{2}\Delta E_{\text{pot}}. \tag{5.38}$$

Half of the potential energy released during the contraction must be converted into kinetic energy of the particles. Since the total number of particles does not change after the fusion processes have finished, the temperature must increase. The residual half can be either radiated or transferred into potential energy. The first possibility results in the increase of the luminosity, the second leads to the increase of the star radius R. For stars with masses $M > 0.25 M_{\odot}$ the enlargement of the star size can be so large, that the luminosity (5.26) increases although the surface temperature decreases. The star becomes a red giant, for very massive stars even a super-giant. The star moves in the HRD to the upper right. On the way from the mains sequence it passes through several stages of instability, which result in changes of size, temperature and luminosity. This is indicated in the HRD by oscillations of its path (see Fig. 5.33).

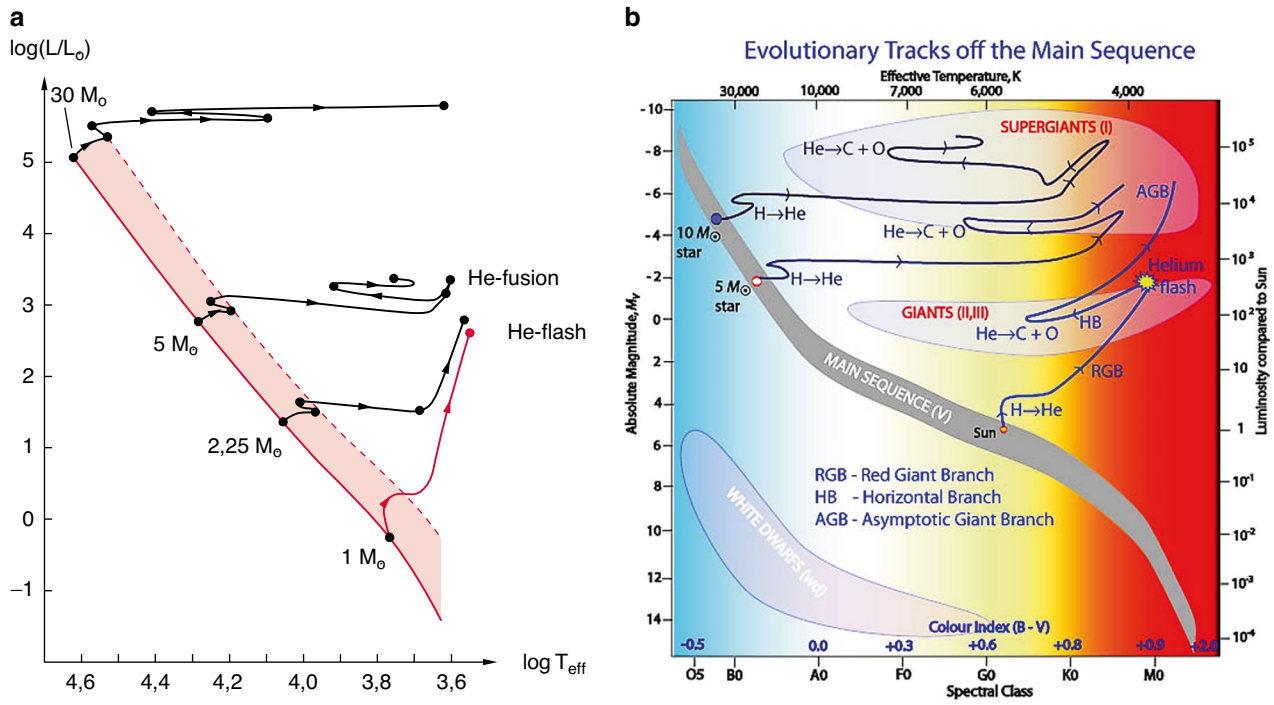


Fig. 5.33 **a** Development paths in the HRD during the time after having left the main sequence for stars with different masses (from Karttunen et al. Astronomy Springer Heidelberg 1994). **b** The path of stars in the HRD, after leaving the main sequence (Australian Telescope)

5.6.1 Development of Stars with small Masses

The development of stars after leaving the main sequence is most simple for small stars with masses $0.08\,M_\odot < M < 0.25\,M_\odot$. Since they had been fully convective during their stable burning phase, they are able to burn nearly their total hydrogen content. A the end of their stable burning phase they consist mainly of helium. However, they never reach sufficiently high temperatures to burn helium to carbon. Therefore they slowly cool down after their hydrogen burning phase ends, and contract. The total collapse is prevented, because with decreasing volume the kinetic energy of the electrons increase (see the quantum–mechanical explanation in Sect. 5.7) and therefore also their pressure, which compensates the gravitational pressure and prevents a further collapse. They become finally **white dwarfs**.

5.6.2 Development of Stars with Medium Masses

For stars with masses $0.25\,M_\odot \leq M \leq 2.5\,M_\odot$ the development, after leaving the main sequence, proceeds more complex. After the end of hydrogen burning in the core the central part (core plus surrounding layers) contracts and heats up because the gravitational energy is converted into kinetic energy of the particles. The layers surrounding the core achieve sufficient high temperature to enable fusion processes and hydrogen burning starts in the layers around the core. This causes the increase of their temperature and the outer layers of the star expand. The star becomes a red giant.

For stars with $M > 2\,M_\odot$ the central temperature reaches values of 10^8 K which is higher than the ignition temperature of the 3α-process, where three α-particles (He-nuclei) fuse to form a carbon nucleus according to the reactions

$$^4\mathrm{He} + ^4\mathrm{He} \rightarrow ^8\mathrm{Be} + \gamma \qquad (5.39a)$$

$$(\Delta E = -0.1\mathrm{MeV}),$$

$$^8\mathrm{Be} + ^4\mathrm{He} \rightarrow ^{12}\mathrm{C} + \gamma \qquad (5.39b)$$

$$(\Delta E = +7.4\mathrm{MeV}).$$

Since the ^{8}Be nucleus is unstable and decays already after $2.5 \cdot 10^{-16}$s, the process (5.39b) must proceed simultaneously with (5.39a), i.e. 3 α-particles must collide simultaneously.

While for stars with masses $M > 1.4\ M_\odot$ with a convective core the intermixing of the material results in a continuous He-burning, this is different for stars with $M < 1.4\ M_\odot$, e.g. our sun. Here the temperature rise at the contraction of the central area is so steep, that He- fusion starts explosively.

The reason for this situation is the following: For these stars the density of the N electrons in the central part with volume V is determined by the Fermi- statistics. Since for each electron the volume V/N is available, its minimum momentum p is determined by the quantum–mechanical uncertainty relation, $p^2 \geq \hbar^2$ /(V/N).This results in a minimum electron pressure p_e which is called "**degenerate pressure**" because it does not depend on the temperature and exceeds the classical pressure $p_i = n_i \cdot k \cdot T$ of the ions with $n_i = N/V$, if the volume V for the hot plasma (electrons + ions) is small enough.

The total pressure $p_e + p_i$ of electrons and ions just compensates the gravitational pressure. The increase of the thermal pressure due to the temperature rise is insignificant as long as $p_i < p_e$. Therefore the energy produced by the 3α-fusion cannot be converted into expansion energy but only into kinetic energy of the particles. This leads to a steep increase of the temperature until the thermal pressure exceeds the degenerate pressure. Now the core can expands and its temperature decreases. This slows down the fusion rate and therefore the energy production rate. Now the normal continuous He-burning takes place.

Because of the large temperature rise during the stage with $p_i < p_e$ the energy production rate in the core is higher than the energy transport to the surface. The star is not in a stable equilibrium. The radiation from the core and the explosive pressure shock wave are absorbed in the outer layers of the star, which are therefore heated up. The luminosity of the star increases steeply and reaches values up to 10^4–10^6 times the normal value (**Helium flash**). However, the star is not teared up by the shock wave because the energy is absorbed by the outer layers not by the inner part of the star.

After the Helium flash the star passes through several unstable phases, indicated in the HRD of Fig. 5.33a by corresponding irregular oscillations. It reaches finally the horizontal branch of the red giants.

The radius of such a red giant can grow up to $250R_\odot$. This implies that our sun at the end of its stable burning phase (in about 6–8 Billion years) will extend beyond the earth orbit. The earth will then be located inside the sun and will therefore probably evaporate. This means that life on earth has a natural end unless it will be extinguished much earlier by human wrongdoing or by some catastrophic cosmic events.

When the helium supply in the core is exhausted, there are still the two shells of Fig. 5.30, where fusion takes place: The inner shell with He-burning and the outer shell with H-burning (Figs. 5.34b and 5.35). Since the energy production in the core has ended, core temperature and pressure decrease and the core contracts while the outer layers

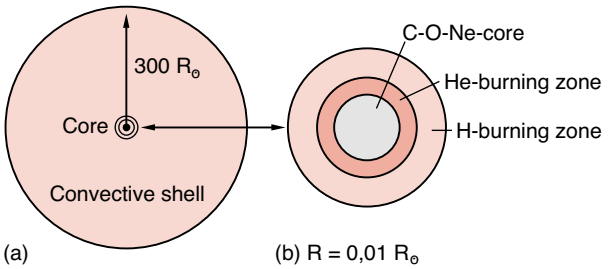


Fig. 5.34 **a** Schematic representation of a red giant star with a core containing C, O and Ne and two inner burning shells **b** magnified drawing of the central part

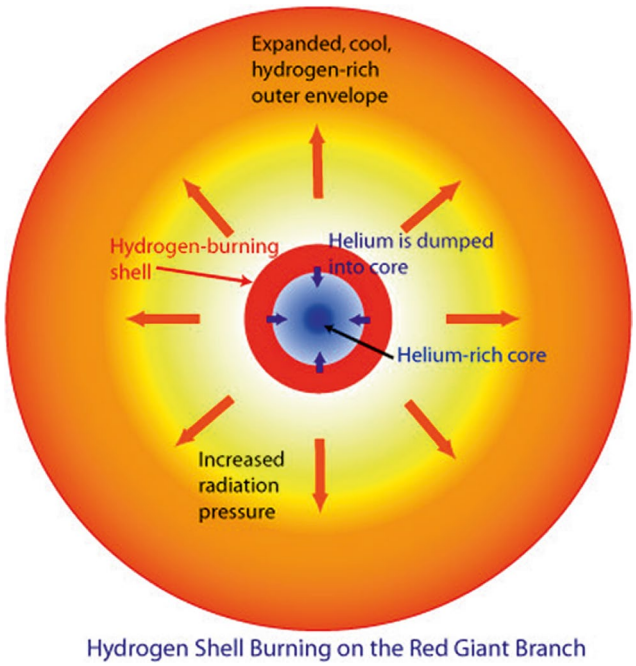

Expanded, cool, hydrogen-rich outer envelope
Hydrogen-burning shell
Helium is dumped into core
Helium-rich core
Increased radiation pressure
Hydrogen Shell Burning on the Red Giant Branch

Fig. 5.35 Hydrogen Burning in the red giant phase of a star with medium mass. (https://www.bing.com/images/search?q=nuclear+fusion+r)

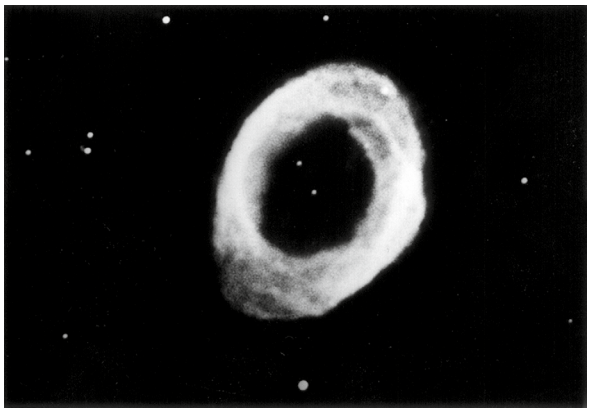


Fig. 5.36 Ring nebula in the constellation Lyra as example of an interplanetary nebula

Fig. 5.37 Helix nebula (Hubble Helix team)

expand. This is an unstable situation and the star ejects part of its envelope, which becomes visible as planetary nebula (Figs. 5.36 and 5.37). The core of the star becomes a **White Dwarf**, if its mass is below 1.4 $M_\odot$ (see Sect. 5.7).

In Fig. 5.33b the paths of stars with masses between 1 and 10 $M_\odot$ is illustrated and in Fig. 5.38 a more detailed picture for the path of a star with 1 $M_\odot$ in the HRD illustrates the future destiny of our sun, which finally becomes after the stage of a red giant a white dwarf.

5.6.3 The Development of Stars with large Masses and the Synthesis of Elements Up to Iron

In massive stars the hydrogen fusion in the core proceeds similar to that in stars with small masses (see Sect. 5.4.x) The main differences beginn at the He-burning stage, which can start already due to the higher temperature, before the core contracts. The matter in the core is not yet degenerate and the He-burning does not proceed explosively as in stars with smaller masses. There is no helium flash. While in the core Helium fuses to carbon, the hydrogen burning proceeds in the layers adjacent to the core. If sufficient carbon has been produced in the reactions (5.39) new processes of He-buring can start.

When all of the He-supply in the core is exhausted the carbon core starts to contract. The resulting higher temperature ignites the He-reactions (5.39) in the shells adjacent to the core. They have sufficient helium concentrations supplied by the hydrogen fusion. This further increases the

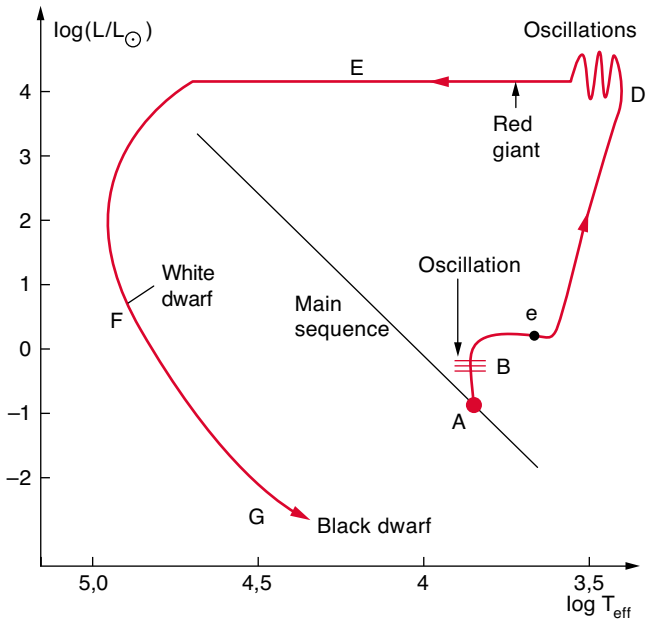


Fig. 5.38 Path in the HRD of a star with $M = M_{\odot}$. **A**: stable hydrogen burning on the main sequence. **B**: end of H-burning, **D**: helium flash, **E**: contraction with internal heating, **F**: white dwarf, **G**: cooling

temperature and enables new fusion processes for stars with $M > 8\,M_{\odot}$, as e.g.

– C-Burning

$$^{12}\mathrm{C} + ^{4}\mathrm{He} \rightarrow ^{16}\mathrm{O} + \gamma \qquad (5.40a)$$

$$(\Delta E = +7.15\mathrm{MeV}),$$

$$^{16}\mathrm{O} + ^{4}\mathrm{He} \rightarrow ^{20}\mathrm{Ne} + \gamma \qquad (5.40b)$$

$$(\Delta E = +4.73\mathrm{MeV}),$$

where oxygen and Neon are produced. With decreasing He-concentration the processes (5.39) decrease, because their probability is proportional to the third power of the He-concentration $\left(\propto N_{\mathrm{He}}^{3}\right)$ due to the 3 α-process. Now the process (5.40a) supplies the major part of the fusion energy. At the end of the central He-burning the core contains about 80% oxygen O, 10% carbon C and 10% neon Ne. Now 2 carbon nuclei can fuse to Magnesium Mg and sodium Na, according to the scheme

– Ne-Burning

$$^{12}\mathrm{C} + ^{12}\mathrm{C} \begin{cases} \rightarrow ^{24}\mathrm{Mg} + \gamma \\ \rightarrow ^{23}\mathrm{Mg} + \mathrm{n} \\ \rightarrow ^{23}\mathrm{Na} + \mathrm{p} \\ (\Delta E = 2.2\mathrm{MeV}) \\ \rightarrow ^{20}\mathrm{Ne} + ^{4}\mathrm{He} \\ (\Delta E = 4.6\mathrm{MeV}), \end{cases} \qquad (5.40c)$$

At temperatures above 10^{9}K photo-dissociation of the neon nuclei starts

$$^{20}\mathrm{Ne} + \gamma \rightarrow ^{16}\mathrm{O} + ^{4}\mathrm{He}, \qquad (5.41a)$$

where He-nuclei are produced. This opens the way for the reactions

$$\begin{aligned} ^{16}\mathrm{O} + ^{4}\mathrm{He} &\rightarrow ^{20}\mathrm{Ne}, \\ ^{20}\mathrm{Ne} + ^{4}\mathrm{He} &\rightarrow ^{24}\mathrm{Mg} + \gamma, \\ ^{24}\mathrm{Mg} + ^{4}\mathrm{He} &\rightarrow ^{28}\mathrm{Si} + \gamma, \end{aligned} \qquad (5.41b)$$

– O-Burning

$$^{16}\mathrm{O} + ^{16}\mathrm{O} \begin{cases} \rightarrow ^{32}\mathrm{S} + \gamma \\ \rightarrow ^{31}\mathrm{S} + \mathrm{n} \\ \rightarrow ^{31}\mathrm{P} + \mathrm{p} \\ \rightarrow ^{28}\mathrm{Si} + ^{4}\mathrm{He}, \end{cases} \qquad (5.42)$$

– Si-Burning

Before such high temperatures are reached that are necessary to fuse two Si-nuclei, photodissociation of the Si-nuclei would start. Therefore the Si-burning is a very complex network of numerous photo-dissociation processes, capture reactions of α-particles resulting in the production of nickel nuclei and iron nuclei

$$\begin{aligned} ^{28}\mathrm{Si} + ^{28}\mathrm{Si} &\rightarrow ^{56}\mathrm{Ni} + \gamma \\ &\rightarrow ^{52}\mathrm{Fe} + ^{4}\mathrm{He}. \end{aligned} \qquad (5.43)$$

These fusion processes alternate, where the central temperature increases at each fusion process.

The burning times decrease with higher temperatures. For example for a star with $M = 25\ M_{\odot}$ the oxygen burning lasts for abojut 180 d, the silicon burning only a few days. In Table 5.7 temperatures, densities in the core and burning times for the different fusion processes for a star with $M = 18\ M_{\odot}$ and $R = 50\ R_{\odot}$ are compiled. This illustrates that with each subsequent burning phase the temperature becomes higher and higher and the burning duration shorter and shorter.

For stars in the lower range of massive stars the synthesis of elements ends with the reaction (5.40c), because the central temperature is not high enough to fuse heavier elements with a higher Coulomb barrier.

The energy released by the reactions (5.40a)–(5.43) increases the temperature in the different shells around the core. This shifts the burning shells farther to the surface of the star.

Up to the production of iron by fusion processes, energy is gained at each fusion step because the binding energy per nucleon has its maximum around the iron nucleus (Fig. 5.39). In Fig. 5.40 the region around the most stable nuclei is shown on an expanded scale.

Table. 5.7 The different fusion processes and their burning time in a star with mass $M = 18M_{\odot}$

Fusion elements	Nucleo synthesis	Temperature in 10^6 K	Density (kg/cm^3)	Burning time
H	Hydrogen burning	40	0.006	10 Mio. years
He	Helium burning	190	1,1	1 Mio. years
C	Carbon burning	740	240	10.000 years
Ne	Neon burning	1600	7400	10 years
O	Oxygen burning	2100	16.000	5 years
Si	Silicon burning	3400	50.000	1 week
Fe_core	Nuclear fusion of heavy elements	10.000	10.000.000	–

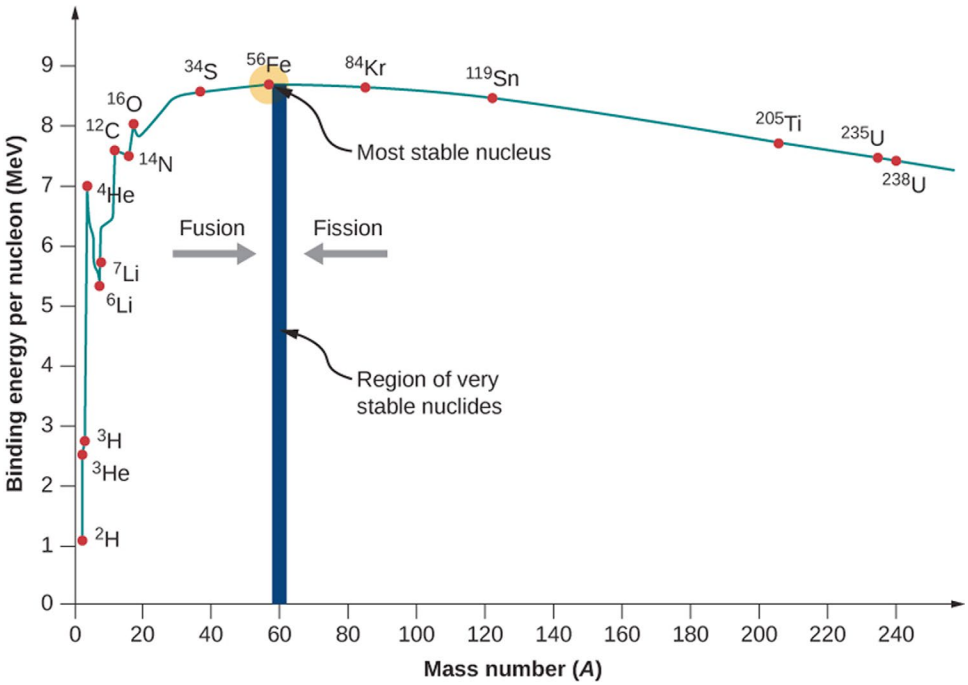


Fig. 5.39 Binding energy per nucleon for the different nuclei with its maximum around the iron nucleus (https://opentextbc.ca/universityphysicsv3openstax/chapter)

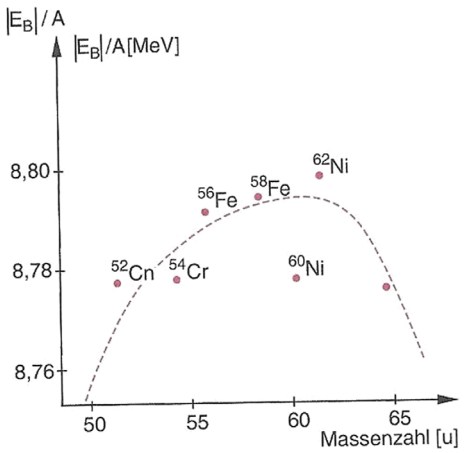


Fig. 5.40 Nuclear binding- energy around the most stable nuclei

Such a massive star has then an "*onion shell-structure*" shown in Fig. 5.41, with a core consisting mainly of iron and several shells where still fusion processes go on. The outer envelope contains mainly hydrogen, but it is too cold to ignite fusion processes.

These fusion processes in the different shells, which produce energy, results in an expansion of the star. It becomes a super giant (Fig. 5.33b) with a very dense core of iron.

Because of the explosive energy release during the Si-burning the temperature steeply rises but stays still below the Fermi-temperature of the electrons, which is very high during the synthesis of the heavier elements and the resulting high density. The Fermi-energy is about 1–10 meV per electron (depending on the mass of the star). Now it becomes energetically favourable that the electrons recombine with protons in the inverse ß-decay

$$e^- + p \rightarrow n + \nu_e \qquad (5.44)$$

and produce neutrons and neutrinos.

The sudden end of the Si-burning finishes the energy production in the core. The thermal pressure decreases fast and the core collapses, due to the gravitational pressure.

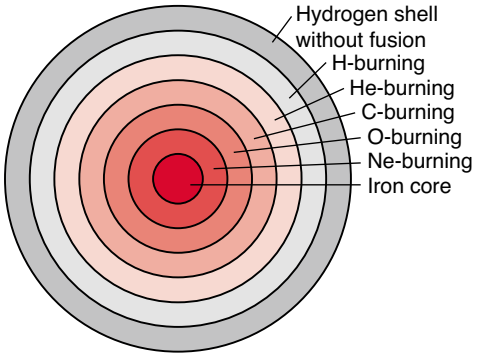


Fig. 5.41 Onion-shell structure of a massive star after the extinction of the last fusion source in the core. (After H Scheffler, H. Elsässer: Physik der Sterne und der Sonne, Spektrum Verlag Heidelberg 1990)

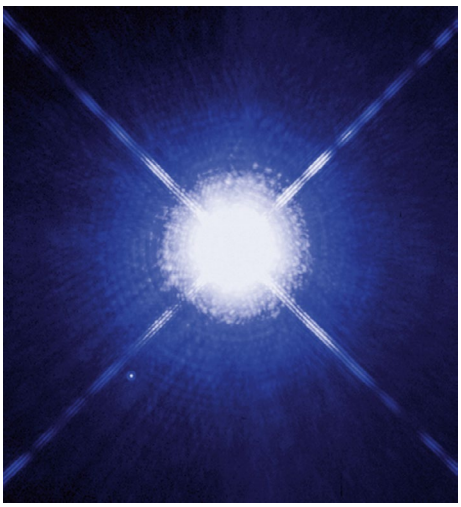


Fig. 5.42 Hubble-Photo of Sirius A (bright star) and for the white dwarf Sirius B (small point on the lower left) (NASA)

The star material precipitates in a free fall towards the center until the density becomes so high that the gravitational pressure is compensated by the degenerate pressure of the neutrons (see Sect. 5.7.1). This happens at a core radius of about 10 km (!!).

This collapse of a star with mass $M = 8\,M_\odot$ from a radius $R_1 = 10^5$ km to a final radius $R_2 = 10$ km releases the immense gravitational energy

$$\Delta E_G = G \cdot M^2 \left(\frac{1}{R_2} - \frac{1}{R_1} \right) \approx 2 \cdot 10^{48} \text{J}. \qquad (5.45)$$

The collapse is suddenly stopped when the density of the core reaches the nuclear density ($\varrho = 10^{16} \text{kg/m}^3$). This sudden stop generates a recoil pressure wave which travels with the initial velocity $v = 5 \cdot 10^4 \text{km/s}$ in the radial directions outwards. It pushes the outer hydrogen-rich envelope

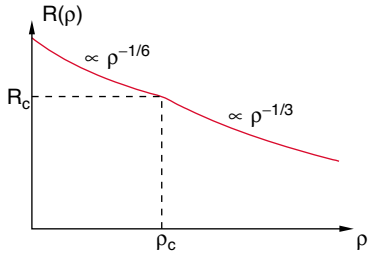


Fig. 5.43 Radius of a white dwarf as a function of the mass density ϱ

of the star explosively out of the star. The released energy is mainly emitted as electromagnetic and gravitational waves and as neutrinos where the neutrinos take the by far major part of the total energy. This phenomenon is called **Supernova-explosion of type II** (see Sect. 5.9).

The remaining core is a Neutron Star (see Sect. 5.7.3). The energy released by radiation amounts to about $2 \cdot 10^{44}$J, which is less than 0.01% of the gravitational energy. Part of the energy is consumed by photo-dissociation of heavy nuclei, produced during the fusion period in the shells. The major part is, however, emitted as neutrinos, which have been produced by the reaction (5.44).

For very massive stars ($M > 8\,M_\odot$) the collaps of the central region can no longer be stopped by the degenerate pressure of the neutrons, because the gravitational pressure is too high. In this case a black hole is created which represents the end-stage of the star (see Sect. 5.7.4).

Before we shall discuss the different final states of stars we have to deal with the characteristic features of degenerate matter.

5.7 End-Stages of Stars

After the nuclear fusion processes in stars has finished, stars can end up their live either as white dwarfs ($M < 1.4\,M_\odot$), neutron stars ($M \approx 8\,M_\odot$) or black holes ($M > 8\,M_\odot$). The matter in white dwarfs and in neutron stars does not obey classical thermo-dynamics but must be described by quantum–mechanical models. The matter is called "*degenerate*".

5.7.1 Degenerate Star-Material

For stars on the main sequence in the HRD (e.g. our sun) the matter in the inner part of the star can be regarded in a good approximation as an ideal gas (see Sect. 4.x) The equation of state is then

$$p \cdot V_M = N_A \cdot k \cdot T. \qquad (5.46)$$

Replacing the mole-volume

$$V_M = N_A \cdot \overline{\mu} \cdot m_H/\varrho \tag{5.47}$$

by the Avogadro constant N_A and the mean mass $\overline{\mu} \cdot m_H$ of the particles (the quantity μ gives the mean mass of the particles in units of the hydrogen mass), the ratio of pressure p and mass density ϱ can be written as

$$\frac{p}{\varrho} = \frac{k}{\overline{\mu} \cdot m_H} \cdot T. \tag{5.48}$$

▶ The ratio of pressure p and mass density ϱ is for nondegenerate matter proportional to the temperature T.

Division by the square c^2 of the velocity of light c gives:

$$\frac{p}{\varrho c^2} = \frac{kT}{\overline{\mu} \cdot m_H c^2}. \tag{5.49}$$

The right side of (5.49) is the ratio of thermal energy $k \cdot T$ and mass energy $\overline{\mu} \cdot m_H c^2$ of a particle with the mean mass $\overline{\mu} \cdot m_H$. If the thermal energy $k \cdot T$ is larger than the ionization energy, the gas becomes ionized and the mean mass is instead of $\overline{\mu} \cdot m_H$ given by

$$\overline{\mu} = \frac{1}{2}\left(m_p + m_e\right)/m_p \approx 1/2. \tag{5.50}$$

Example

For typical mean temperatures $T = 10^7$K in the central part of the star we get $k \cdot T \approx 10^{-16}$J = 1keV. The mass energy of protons ($\overline{\mu} = 1$) is $m_p c^2 \approx 1$GeV. The ratio of thermal to mass energy is therefore $kT/\left(m_p c^2\right) \approx 10^{-6}$. ◄

Matter under conditions where the thermal energy is small compared to the mass energy is called non-degenerate.

5.7.2 Equation of State for Degenerate Matter

If the gravitational pressure on the inner part of a star exceeds the thermal pressure p, the gas collapses and the density increases until the quantum mechanical Fermi pressure of the electrons can compensate the gravitational pressure. The quantum mechanical uncertainty relation states that the product of the minimum uncertainty of the momentum component Δp_x and the spatial uncertainty Δx follows the relation

$$\Delta p_x \cdot \Delta x \geq \hbar. \tag{5.51}$$

If electrons are confined in a volume $V = d^3$, Eq. (5.51) is valid for each of the three components. For the absolute value of the momentum, which must be larger than its uncertainty, we obtain

$$p^2 = p_F^2 = p_x^2 + p_y^2 + p_z^2 \geq 3\hbar^2/d^2. \tag{5.52}$$

The quantity p_F is called the Fermi-momentum.

The Fermi-Energy is then

$$E_F = \frac{p_F^2}{2m_e} \geq \frac{3\hbar^2}{2m_e \cdot d^2} \approx \frac{\hbar^2}{m_e d^2}. \tag{5.53}$$

The smaller the volume is in which the electrons are confined the larger is the Fermi-energy.

If for a gas confined in the volume V the Fermi energy is $E_F \geq kT$, the gas is called *degenerate.*

Note: In the following estimations only the magnitude of the effects is regarded while factors of the order of magnitude 1 are neglected.

The equation of state for a degenerate gas can be obtained if in (5.49) the thermal energy kT is replaced by the Fermi energy E_F. This gives

$$\frac{p}{\varrho c^2} \approx \frac{E_F}{\overline{\mu} \cdot m_p c^2} \approx \frac{\hbar^2}{\overline{\mu} \cdot m_p \cdot m_e c^2 \cdot d^2}, \tag{5.54}$$

where $\overline{\mu} m_p$ is the mean mass of all particles and $m_e \ll \overline{\mu} \cdot m_p$ the mass of an electron. For $E_F \gg kT$ the thermal pressure p of the gas particles is very small compared to the Fermi-pressure (degenerate pressure) of the electrons. On the other side is the rest mass energy $m_p c^2$ of the protons very much larger than that of the electrons. The mass density

$$\varrho = \frac{m_e + \overline{\mu} m_p}{d^3} \approx \frac{\overline{\mu} m_p}{d^3} \tag{5.55}$$

is essentially given by the nuclei, but much less by the electrons, because ($m_e \ll \overline{\mu} m_p$). Solving (5.55) for the volume with dimensions d and inserting the result into (5.54) we obtain

$$\frac{p}{\varrho c^2} \approx \frac{\hbar^2}{m_e c^2 \cdot \overline{\mu} m_p} \left(\frac{\varrho}{\overline{\mu} \cdot m_p}\right)^{2/3}. \tag{5.56}$$

Introducing the critical density

$$\begin{aligned} \varrho_c &= \frac{\overline{\mu} \cdot m_p}{(\hbar/m_e c)^3} \\ &\approx 3 \cdot 10^{10} \text{kg/m}^3 \quad \text{for} \quad \overline{\mu} = 1 \end{aligned} \tag{5.57}$$

where the mean distance between the protons (and therefore in the neutral plasma also between the electrons) is equal to the Compton wavelength λ_C of the electrons

$$\lambda_c = \frac{h}{m_e \cdot c} = 2 \cdot 10^{-12} m \tag{5.58}$$

the equation of state (5.56) can be written as

$$\frac{p}{\varrho c^2} = \frac{m_e}{\overline{\mu} \cdot m_p} \cdot \left(\frac{\varrho}{\varrho_c}\right)^{2/3} \quad \text{for} \quad \varrho < \varrho_c. \tag{5.59}$$

- For the critical density $\varrho = \varrho_c$ the ratio of Fermi pressure of the electrons and mass energy density ϱc^2 (contributed by the atomic nuclei) is proportional to the ratio of electron mass m_e and mean nuclear mass $\overline{\mu} \cdot m_p$.

The significance of the critical density can be illustrated by the following consideration: Inserting into (5.51) for Δx the Compton wavelength $\lambda_C = h/(m_e \cdot c)$ for electrons, one obtains

$$\frac{p_F \cdot h}{m_e \cdot c} \gtrapprox h \Rightarrow p_F \gtrapprox m_e \cdot c. \tag{5.60}$$

this means: If the electrons are included into a volume V where for each electron only the volume λ_c^3 is available, the velocity of the electrons approaches the velocity c of light and their total energy becomes large compared to their rest mass energy due to their Fermi-momentum.

For $\varrho > \varrho_c$ the electron energy increases further (their relativistic mass increases) and we can insert in (5.54) for the Fermi energy the relation

$$E_F = p_F \cdot c$$

This gives the equation of state

$$\frac{p}{\varrho c^2} \approx \frac{m_e}{\overline{\mu} m_p} \left(\frac{\varrho}{\varrho_c}\right)^{1/3} \quad \text{for} \quad \varrho > \varrho_c. \tag{5.61}$$

The general equation of state for degenerate gases, where the Fermi pressure is large compared to the thermal pressure, then becomes.

$$\frac{p}{\varrho c^2} \approx \frac{m_e}{\overline{\mu} m_p} \left(\frac{\varrho}{\varrho_c}\right)^{n/3} \tag{5.62}$$

With.

$n = 2$ for $\varrho < \varrho_c$ and $n = 1$ for $\varrho > \varrho_c$.

This result obtained by estimations agrees well with exact calculations in the mass density range $10^4 kg/m^3 < \varrho < 10^{13} kg/m^3$.

- For degenerate matter pressure and mass density are no longer proportional.

For $\varrho < \varrho_c$ is $p \propto \varrho^{5/3}$ while for $\varrho > \varrho_c$ is $p \propto \varrho^{4/3}$.

5.7.3 White Dwarfs

We had discussed in Sect. 5.6.2 that stars with masses $M < 2.5 M_\odot$ after the hydrogen burning in the core has ended, collapse and reach such a high density, that the degenerate pressure (Fermi-pressure) of the electrons becomes larger than the thermal pressure of the ions in the hot plasma. The interior of the star consists in this stage, after having left the main sequence in the HRD, of the degenerate electron gas and Helium nuclei. The energy production at the start of the helium burning is therefore not used for expanding the core because the gas pressure, increasing with the increasing temperature remains still below the degenerate electron pressure. Therefore the He-burning proceeds explosively (**helium flash**). However, the central temperature does not exceed the ignition temperature for the carbon burning. The layers around the central part cool down after the end of the He-burning and contract. This contraction releases energy which is transported to the surface of the star and results in a temperature increase of the star envelope which blows up. The star becomes a red giant. Because of its large surface the energy loss caused by the emitted radiation cools the outer parts of the star down. The particles in the outer layers have still enough kinetic energy to be ejected from the surface. There is a continuous gas flow out of the star which forms a planetary nebula (Fig. 5.37). The star looses by the outgoing gas flow its outer layers while the inner layers contract. They form the core of the star which has, due to the contraction, a very high density and a high temperature. Because there are no longer nuclear relations, the core cools slowly down because it looses energy by emitting radiation. This final stage is called "**White Dwarf**". Its position in the HRD is in the lower left side of the main sequence (Figs. 5.16; 5.33b and 5.38).

White Dwarfs represent the final stage of stars with masses $M < 1.4 M_0$ (after ejection of the star envelope and contraction of the central part to an earth-like size) with radii $R \approx 10^4 km$.

One of the earliest observed examples of a white dwarf is the component B of the double star system Sirius A+B (Fig. 5.42).

Its surface temperature is 25.000 K and therefore about 4.4 times as high as the temperature of the sun surface. Its luminosity is, however, only about 0.3% of the sun

luminosity $L_\odot$. Since $LR^2 \cdot T^4$, the surface of sirius B must be very small. From the measured luminosity and the surface temperature one can deduce the radius of Sirius B from the relation

$$L = \pi R^2 \sigma T^4 \rightarrow R = \left(L/4\pi\sigma T^4\right)^{1/2}$$

The comparison with the corresponding values of our sun gives

$$R = R_\odot \cdot \frac{T_\odot^2}{T^2} \cdot \sqrt{L/L_\odot} \tag{5.63}$$

Inserting the known values of the sun yields

$$6 \cdot 10^{-2} R_\odot \approx 5 \cdot 10^3 \text{km}.$$

▶ Meanwhile a large number of white dwarfs has been discovered. They contribute an essential part of all stars in our galaxy.

The mass of white dwarfs is about $M \approx 1\,M_\odot$, their size is, however, similar to that of the earth. The mass density has therefore the eminent large value $\varrho \approx 10^9 \text{kg/m}^3$.

1cm^3 of the matter of a white dwarf would have on earth the weight of 1000 kg(!).

In order to determine the relation between mass, radius and mass density of white dwarfs we start with the equation of state (5.62). For a stable white dwarf, the degenerate pressure must be equal to the gravitational pressure (5.17)

$$p_\text{G} = -\frac{3}{8}\frac{M^2 \cdot G}{\pi R^4}. \tag{5.64}$$

Inserting $\varrho = M/\left(\left(\frac{4}{3}\right)\pi R^3\right)$ we obtain with (5.59)

$$\begin{aligned} \frac{p}{\varrho c^2} &= \frac{1}{2}\frac{G \cdot M}{Rc^2} \\ &= \frac{m_\text{e}}{\mu m_\text{p}}\left(\frac{\varrho}{\varrho_\text{c}}\right)^{n/3}. \end{aligned} \tag{5.65}$$

The same result has been already derived in Sect. 5.7.1, based on general reflections for degenerate matter. Replacing the radius R by the mass

$$M = \left(\frac{4}{3}\right)\pi\varrho R^3 \Rightarrow R = \left(\frac{3M}{4\pi\varrho}\right)^{1/3}$$

We can write (5.65) as

$$\frac{p}{\varrho c^2} \approx \left(\frac{\pi}{6}\right)^{1/3}\frac{G}{c^2}M^{2/3}\varrho^{1/3}. \tag{5.66}$$

Inserting this expression into (5.65) gives for the masses of white dwarfs, depending on their mass density ϱ

$$M(\varrho) = \sqrt{6/\pi} \cdot \left(\frac{m_\text{e}c^2}{G \cdot \mu m_\text{p}}\right)^{3/2}\frac{\sqrt{\varrho}}{\varrho_\text{c}} \quad \text{for} \quad \varrho < \varrho_\text{c}, \tag{5.67a}$$

$$M(\varrho) = \sqrt{6/\pi} \cdot \left(\frac{m_\text{e}c^2}{G \cdot \mu m_\text{p}}\right)^{3/2}\frac{1}{\sqrt{\varrho_\text{c}}} \quad \text{for} \quad \varrho > \varrho_\text{c}. \tag{5.67b}$$

This is a remarkable result: The Mass $M(\varrho)$ increases with the square root of the density ϱ as long as $\varrho < \varrho_\text{c}$. At the critical density ϱ_c the mass of a stable white dwarf has its upper limit $M_\text{c} = M(\varrho_\text{c})$.

This critical mass M_c is called after its discoverer the Indian-American astrophysicist *Subrahmanyan Chandrasekhar* the **Chandrasekhar Mass Limit**. If the mass further increases, the star collapses and becomes a Neutron Star (see next section).

There are no white dwarfs with masses $M > M_\text{c}$.

Inserting the value of ϱ_c from (5.57), the critical mass M_c can be expressed as

$$M_\text{c}(\varrho_\text{c}) = \sqrt{6/\pi}\left(\frac{\hbar c}{G\left(\mu m_\text{p}\right)^2}\right)^{3/2} \cdot \overline{\mu} m_\text{p}. \tag{5.68}$$

The dimensionless fraction

$$\alpha_\text{G} = \frac{G\left(\mu m_\text{p}\right)^2}{\hbar c} \approx 6 \cdot 10^{-39} \tag{5.69}$$

is called **the finestructure constant of gravitation**.

$$M_\text{c} \approx 1.4\frac{\overline{\mu} m_\text{p}}{\alpha_\text{G}^{3/2}} \tag{5.70}$$

Inserting into (5.68) the numerical values for the constants and taking the average mass of the H–He mixture in the white dwarf as $\overline{\mu} \cdot m_\text{p} = 2.4 m_\text{p}$ we obtain the Chandrasekhar mass limit.

$$M_\text{c} = 3 \cdot 10^{30}\ \text{kg} \approx 1.45\,M_\odot. \tag{5.71}$$

Note: that this mass limit is valid for the residual mass of a star that has passed into its final stadium of a white dwarf. Since the star has, before this transition, ejected its envelope, which might contain a substantial fraction of its total mass, the mass of the white dwarf is considerable smaller than the original mass of the star. Furthermore the mass decreases during the contraction of the core. If a star contracts from a radius R_0 to a much smaller radius $R \ll R_0$ it looses the mass

$$\Delta M = -(\frac{3}{5})GM_0^2\left(\frac{1}{R} - \frac{1}{R_0}\right) \tag{5.72}$$

which is converted into gravitational energy. Therefore the mass of the original star can be quite larger than $1.5\,M_\odot$.

We have so far assumed that the star does not rotate, which is generally not true, since many stars rotate before they contract. Since the angular momentum $I = \theta \cdot \omega$ must be preserved during the collaps the angular velocity ω must increase, because the moment of inertia θ decreases during the collaps. The centrifugal force in a rotating star partly compensates the gravitational force. This leads to a larger mass limit which can increase up to $M_c = 3\,M_\odot$.

The radii of white dwarfs can be estimated as follows:

$$\begin{aligned} R &= \left(\frac{3M}{4\pi\varrho}\right)^{1/3} \\ &= \left(\frac{3}{4\pi}\right)^{1/3} \cdot \left(\frac{M_c}{\varrho_c}\right)^{1/3} \cdot \left(\frac{M}{M_c}\right)^{1/3} \cdot \left(\frac{\varrho_c}{\varrho}\right)^{1/3}. \end{aligned} \tag{5.73}$$

Inserting the critical radius

$$R_c = \left(\frac{3}{4\pi} \cdot \frac{M_c}{\varrho_c}\right)^{1/3} \tag{5.74}$$

which is the radius of the most massive white dwarf with $M = M_c$, we obtain with (5.67a) for the mass the relation for $\varrho < \varrho_c$

$$M = M_c \cdot \sqrt{\varrho/\varrho_c} \tag{5.75}$$

and the radius

$$R = \left(\frac{\rho_c}{\rho}\right)^{1/6} \cdot R_c. \tag{5.76}$$

While for $\varrho > \varrho_c$

$$R = \left(\frac{\varrho_c}{\varrho}\right)^{1/3} \cdot R_c. \tag{5.77}$$

Inserting the numerical values for M_c and ϱ_c from (5.68) and (5.57) gives a critical radius $R_c \approx 10^7\text{m} = 10^4\text{km}$.

The radius of white dwarfs lies around 10^4 km, which agrees with the observations (see the example of Sirius B). It is for $\varrho < \varrho_c$ larger than R_c and for $\varrho > \varrho_c$ smaller than R_c (Fig. 5.43).

From (5.75) and (5.76) the important relation between mass and radius of white dwarfs

$$\begin{aligned} M \cdot R^3 &= M_c \cdot R_c^3 \\ \Rightarrow R &= R_c \cdot (M_c/M)^{1/3} \end{aligned} \tag{5.78}$$

follows (Fig. 5.44).

▶ The radii of white dwarfs

$$\begin{aligned} R &= (M_c/M)^{1/3} \cdot R_c \\ &= \text{const.} \cdot M^{-1/3} \end{aligned}$$

decrease with increasing mass! For the Chandrasekhar mass limit $M = M_c$ they have the value $R_c = 10^4$km.

In Fig. 5.44 the dependence of the radius R(M) on the mass M is shown for two different mean masses $\overline{m} = \overline{\mu} \cdot m_p$ of the particles in the white dwarf.

For white dwarfs the ratio of Fermi-pressure and mass energy density

$$\frac{p_F}{\varrho c^2} = \frac{m_e}{\mu m_p}\left(\frac{\varrho}{\varrho_c}\right)^{2/3} \approx 10^{-4} \tag{5.79}$$

is of the order 10^{-4}.

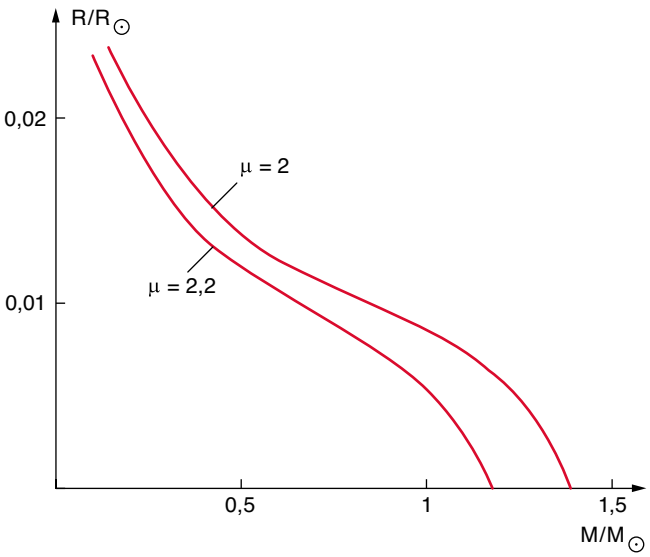


Fig. 5.44 The radii of white dwarfs as a function of their mass (without taking into account the Coulomb interaction) for two different mean masses $\mu = \overline{m/mp}$ of the gas atoms

Note: For the degenerate electron gas in a metal this ratio is 10^{-5}.

The interior of white dwarfs consists mainly of helium, The outer layers contain helium and hydrogen. The surface gravity is about 10^5 times that of the earth.

5.7.4 Neutron Stars

If the density ϱ in the central part of a massive star in its final state after the nuclear fusion processes have ended, exceeds the critical density ϱ_c of a white dwarf the Fermi-energy E_F of the electrons must further increase in order to compensate the gravitational pressure. However, as soon as E_F exceeds the limit

$$E_F^c \geq (m_n - m_p)c^2 \tag{5.80}$$

of the mass energy difference between neutron and proton, electrons can combine with protons to form neutrons

$$e^- + p \rightarrow n + \nu_e \quad (\text{inverse}\,\beta - \text{decay}). \tag{5.81}$$

This decreases the density of electrons in the plasma of electrons and atomic nuclei with the result that the Fermi-pressure of the electrons also decreases. The central part of the star becomes unstable, because the gravitational pressure is no longer compensated by the Fermi-pressure. Therefore the central part collapses until the density has reached the limiting value $\varrho = 10^{16}\text{kg/m}^3$. Now the volume for the neutrons becomes so small that the degenerate pressure of the neutrons increases to such a value that it can compensate the gravitational pressure.

▶ While White Dwarfs are stabilized by the degenerate pressure of the electrons for Neutron stars the stabilization is achieved by the degenerate pressure of the neutrons.

5.7.4.1 Density, Mass and Radius of a Neutron Star

We obtain for neutron stars the corresponding relations for pressure and density as for White Dwarfs in Sect. 5.7.3), if we replace the electron mass m_e by the neutron mass m_n. The ratio of Fermi-pressure p_F to mass energy density ϱc^2 in the neutron star which consists mainly of neutrons is now

$$\frac{p_F}{\varrho c^2} \approx \frac{m_n}{\mu \cdot m_p} \cdot \left(\frac{\varrho}{\varrho_1}\right)^{n/3} \tag{5.82}$$

where the critical density ϱ_{cN} for neutron stars now becomes

$$\varrho_{cN} = \frac{(\mu+1)m_p}{(\hbar/m_n c)^3} \approx 10^{20}\text{kg/m}^3. \tag{5.83}$$

It is larger by 10 orders of magnitude (!) than for white Dwarfs because of the larger neutron mass and the much smaller volume of the neutron star for a comparable total mass M.

For the masses of the neutron stars we obtain, analogous to (5.67), if we again replace the electron mass m_e by the neutron mass m_n for $\varrho < \varrho_{cN}$

$$M(\varrho) = \sqrt{6/\pi} \cdot \left(\frac{m_n c^2}{G \cdot \mu m_p}\right)^{3/2} \frac{\sqrt{\varrho}}{\varrho_1}. \tag{5.84}$$

Here we also get a limiting upper mass as for the white dwarfs, which is reached for $\varrho = \varrho_c$.

$$M_c = M(\varrho_c) \approx 3.2\, M_\odot \tag{5.85}$$

It is called the **Oppenheimer-Volkow limiting Mass**. For $M > M_c$ the degenerate pressure of the neutrons can no longer withstand the gravitational pressure. The star collapses to a black hole (Sect. 5.7.5).

Neutron stars have a limiting mass density ϱ_c which is $3 \cdot 10^9$ times larger than that of white dwarfs. The volume of neutron stars is therefore smaller by this factor and the radius by the factor $\left(3 \cdot 10^9\right)^{1/3} \approx 1.4 \cdot 10^3$.

▶ In spite of their large mass $M \approx 2 - 3\ M_\odot$ neutron stars have a radius of about $R \approx 10$km.

5.7.4.2 Internal Structure of Neutron Stars

In spite of their name not the total volume of neutron stars consists of free neutrons. At the surface the gravitational pressure has its minimum value and the surface and layers closely below the surface consist mainly of iron nuclei and free electrons, which had formed before the collaps of the central part of the star. With decreasing radius r more and more protons in the iron nuclei are transformed into neutrons. Therefore the fraction of neutron-rich iron isotopes increases with decreasing r. At a radial distance of 10 m below the surface the pressure becomes already so large that free neutrons can exist. In a layer about 2 km from the surface free neutrons (superfluid neutrons) and a lattice of iron nuclei exist side by side (Fig. 5.45) where the fraction of iron nuclei decreases with r from 100% (r = R) to 0% (r = 0).

In the inner part of the neutron star only free neutrons and a small fraction of free protons and electrons exist. At temperatures below 10^{11}K the neutrons form a superfluid liquid (Fig. 5.46a) and the protons are superconductors,

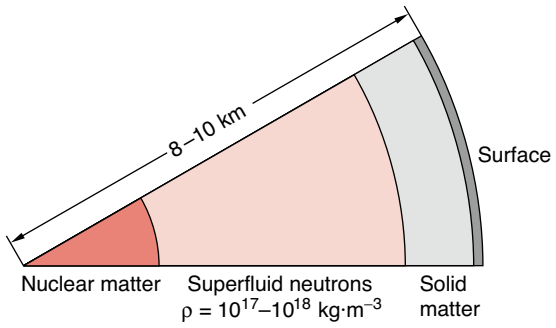


Fig. 5.45 Schematic structure of a neutron star

i.e. they form Cooper pairs of two protons with antiparallel spins according to accepted models of neutron stars.

Up to now it is unknown which state of matter is prevalent in the central part of a neutron star. Here the mass density is larger than that of atomic nuclei. Some people speculate that the most inner part of a neutron star might consist of a quark-gluon plasma.

Plotting the mass of stars in degenerate states against the mass density ϱ, we obtain the curves shown in Fig. 5.47. In those ranges of ϱ where $M(\varrho)$ has a negative slope ($\mathrm{d}M/\mathrm{d}\varrho < 0$) no stable stars exist. They collaps until they reach a range with positive slope.

5.7.4.3 The Formation of Neutron Stars

Neutron stars represent the end phase of stars with sufficiently large masses where at the end of their nuclear fusion processes in the shells only the Silicon-burning in the core continues. For a star with $M = 15\,M_\odot$ this lasts, however, only for one hour. This process therefore proceeds explosively and results in a very large temperature and pressure rise in the core.

Typical parameters at the end of the burning phase are for a star with $M = 15\,M_\odot$.

Mass of the core $M = 1.5\,M_\odot$ which is only 10% of the total mass; Central temperature $T \approx 8 \cdot 10^9\mathrm{K}$; density at the center: $\varrho \approx 4 \cdot 10^9\mathrm{g/cm^3} = 4 \cdot 10^{12}\mathrm{kg/m^3}$.

The Fermi-energy of the electrons is about 5–10 meV and therefore high enough to enable the processes

$$\mathrm{e}^- + \mathrm{p} \rightarrow \mathrm{n} + \overline{\nu}_\mathrm{e}, \tag{5.86a}$$

$$\mathrm{e}^- + {}^Z\mathrm{K} \rightarrow {}^{Z-1}\mathrm{K} + \nu_\mathrm{e} \tag{5.86b}$$

The protons in the atomic nuclei transfer by electron capture into neutrons. This has a fatal consequence for the star: Due to the decrease of the electrons density the Fermi-pressure of the electrons decreases and can no longer compensate the gravitational pressure. The particles plunge nearly in a free fall towards the center.

Note: that only the core of the star takes place in this free fall. The star envelope does not participate.

The collapse is stopped when the density has reached such a value, that the degenerate pressure of the neutrons can compensate the gravitational pressure. The falling material is reflected at the edge of the dense central range of neutrons. This causes a reflected shock wave which travels in the radial outwards direction. The energy of this shockwave depends on the kind of reflection, which can be soft or hard, depending on the magnitude of the density gradient.

The shockwave can dissociate the iron nuclei, falling towards the center, into protons and neutrons. Although the shockwave looses energy by this dissociation process, the strong increase of the number of particles leads to an explosive increase of the pressure and a pressure wave emerges, which can blow away about half of the star mass (**super nova explosion of type I** see Sect. 5.7.4.5) The neutrinos produced by the transfer of protons into neutrons play a significant role, as will be discussed in the following section.

5.7.4.4 The Crucial Role of Neutrinos for the Formation of Neutron Stars

The neutrinos generated by the formation of the core of the neutron star can leave the central zone nearly unimpeded because of their very small absorption by matter. With increasing mass density the scattering rate of neutrinos by the heavy atomic nuclei becomes, in spite of the small scattering cross section, so large, that the diffusion time of the neutrinos through the core of the neutron star becomes longer than the collapse time. The central part of the neutron star gets "optically thick" for the neutrinos, while the outer part of the star with much lower mass density is still transparent for them.

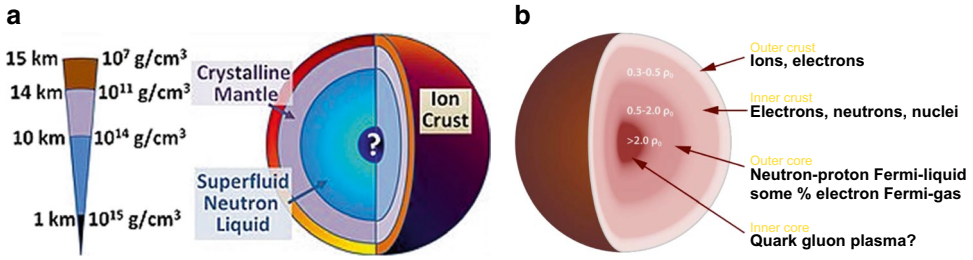


Fig. 5.46 **a** Composition and mass density ρ(R) of a neutron star (Wikipedia) **b** Structure of a neutron star (Robert Schulze, own work.) https://commons.wikimedia.org/w/index.php?cvurid=11364464

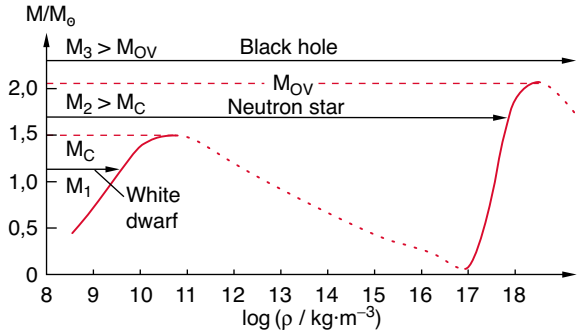


Fig. 5.47 Mass dependence M(ρ) of degenerate stars

The central part of the neutron star now consists of electrons, neutrons, neutrinos and neutron-rich nuclei. The outer part consists of lighter nuclei and electrons.

The border between the "neutrino-opaque region" and the neutrino-transparent region defines the neutrinosphere, quite analogue to the photosphere for the photons (see Sect. 3.x). Initially the neutron star is a hot lepton-rich object, which through the neutrino emission changes over into its final state of a compact star stabilized by the degenerate pressure of the neutrons. In its initial state it might be called a *neutrino-star.* As soon as the neutrinos have left the core they can transmit unimpeded through the envelope of the neutron star because the scattering cross section by the lighter nuclei in the outer parts of the star is much smaller.

The emitted neutrinos play a significant role for the supernova explosion. The breakthrough of the shockwave through the neutrino sphere is supported by the huge amount of neutrinos. In that region of the star, where the shockwave causes a dissociation of the iron nuclei the neutrino scattering rate becomes again smaller causing an increase of the mean free pathlength. Behind the shockwave the neutrinos accumulate and exert an additional pressure onto the shockwave increasing its intensity and enhances the ejection of the shells of the neutron star which leads to the supernova explosion.

5.7.4.5 Energy Release at the Formation of Neutron Stars

The mass of the neutron star is much smaller than that of the original star. This has several reasons:

(a) The envelope of the star has been ejected during the supernova explosion
(b) At the contraction of the central region gravitational energy is released as radiation of photons, neutrinos and gravitational waves.

▶ Neutron stars with $M = 1.4 - 3\,M_\odot$ have only a small fraction of the total mass $M \approx 10 - 15\,M_\odot$ of the original star. The original mass can be therefore much larger than the Oppenheimer-Volkow limiting mass.

Where does the huge energy comes from, which is released as the energy of neutrinos, gravitational waves, optical radiation, -X-rays and the kinetic energy of the ejected gas envelope??

The potential energy of a star with mass M_0 radius R_0 and homogeneous mass density ϱ is (see Problem 5.11)

$$E_{\text{pot}}(R_0) = -\frac{3}{5}G \cdot \frac{M_0^2}{R_0}. \tag{5.87}$$

When the star contracts to the smaller radius $R < R_0$ the potential energy

$$\Delta E_{\text{pot}} = \frac{3}{5}GM_0^2\left(\frac{1}{R} - \frac{1}{R_0}\right) \tag{5.88}$$

is released. For $R \ll R_0$ this is approximately

$$\begin{aligned} M(R) &= M_0(R_0) - \frac{3}{5}G\frac{M_0^2}{Rc^2} \\ \Rightarrow \frac{\Delta M}{M_0} &= -\frac{3}{5}G\frac{M_0}{Rc^2}. \end{aligned} \tag{5.89}$$

An additional mass loss is the star envelope ejected during the supernova explosion.

Example

We assume that the central region of the star which participates in the contraction has a radius $R_0 = 10^5$km and a mass $M_0 = 2\,M_\odot$ which contracts to $R = 10$km; Inserting the numerical values gives:

$$\frac{\Delta M}{M_0} = 0.17,$$

i.e the neutron star looses just by contraction already 17% of its mass. In addition the much larger mass loss by ejection of the envelope contributes to the total mass loss. ◀

The potential energy released by the contraction is enormous:

$$\Delta E_{\text{pot}} = \Delta M \cdot c^2 = 3 \cdot 10^{46}\text{J}.$$

This is about 100 times the energy which our sun radiates during its whole life of 10^{10} years.

Equation (5.88) represents an upper limit for the released potential energy, because the density of the higher star envelopes is much smaller than that of the core. Therefore in reality the released energy will be smaller by at least one order of magnitude.

The total energy released during a supernova explosion is divided between the different kinds of emitted particles and waves:

99% $\approx 3 \cdot 10^{46}$J by neutrinos,

1% $\approx 3 \cdot 10^{44}$J as kinetic energy of the ejected star envelope,

0.01 % as Electro-Magnetic Radiation.

Gravitational waves are only emitted for an asymmetric collapse.

5.7.4.6 Pulsars as Rotating Neutron Stars

In 1967 the PhD student *Jocelyn Bell* (Fig. 5.48) found during a survey of the sky looking for perturbations of radio signals, sharp and very regular radio signals with pulse widths of less than a millisecond and periods between milliseconds and seconds (Fig. 5.49). At first she thought excitedly that these might be signals from other intelligent beings. Telling her thesis adviser, the Astronomy Professor *Anthony Hewish*, he asked her to extend the search of the sources of these signals and it soon became clear that there were many of such "pulsars" and that the signals had too regular pulse sequences which could not carry any messages from aliens.

What was the source of these radio signals?

The pulsar source which had been measured most accurately is located in the crab nebula (Fig. 5.50) which represents the remnant of a supernova explosion observed in the year 1054 by Chinese Astronomers.

Fig. 5.48 Jocelyne Bell (Wikipedia)

The short time duration of the pulses indicates that the source must have a small extension. If the source has the diameter $2R$ the maximum time difference of signals from different points of the source is $\Delta t < 2R/c$.

Example

With the pulse width $\Delta t = 10^{-3}$s $\rightarrow R < 150$km. ◄

More detailed measurements showed that the pulses have a fine structure (Fig. 5.49b), which narrows further down the maximum extension of the source. Since with radio interferometry the pulsars could be accurately localized and their distance determined, it turned out that the pulsars must emit radiation with enormous power.

Such small objects which can emit huge powers must be compact stars with sufficient energy. This leads to neutron stars as possible pulsars (Fig. 5.51).

The observation of variable stars has revealed that radial oscillations caused most of the observed periodic brightness variations. Their period is determined by the time a sound wave needs to traverse the star.

If the observed pulse signals would be produced by radial oscillations, the period must be at least $\Delta t = 2R/v_S \approx 2 \cdot 10^4/2 \cdot 10^3$s ≈ 10s.

The observed millisecond pulses therefore exclude pulsating neutron stars as possible sources.

The correct explanation, which was found only after intense and long standing discussions among astronomers are fast rotating neutron stars as source of the observed pulses. The model, accepted today can be summarized as follows.

During the collapse of a rotating star with mass M, radius R_1, angular velocity ω_1 and angular momentum

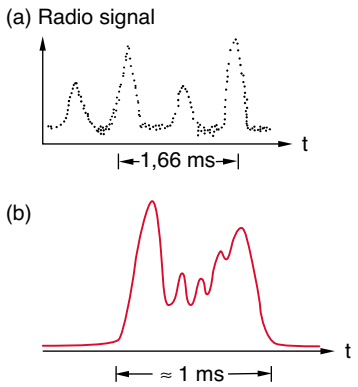


Fig. 5.49 **a** Radio-pulses from the pulsar PSR 1937. **b** fine structure of a single pulse from the pulsar PSR 1237+25

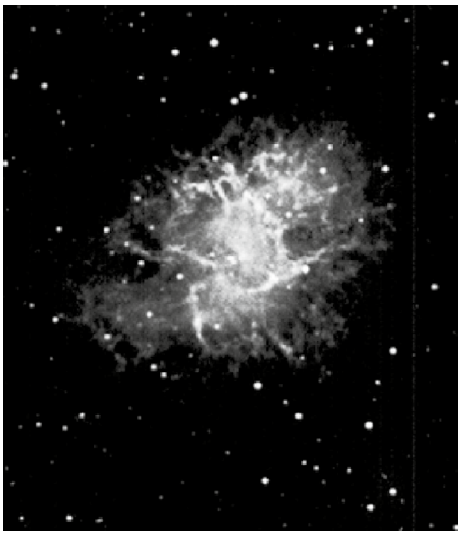


Fig. 5.50 The crab nebula at a distance of 2 kpc is the remnant of a supernova explosion in 1054. In its center a pulsar was detected (with kind permission of the Thüringer Landes-Sternwarte in Tautenburg, Germany)

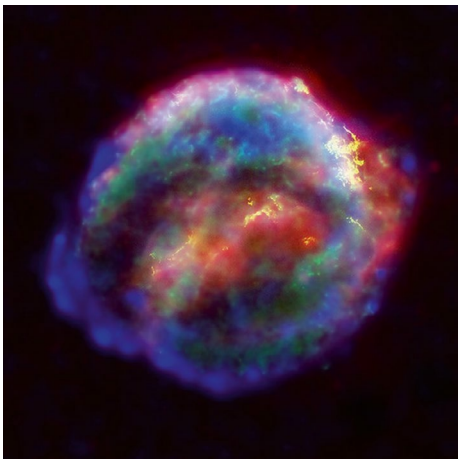


Fig. 5.51 The residue cloud from the supernova observed 1604 in the star constellation Ophiuchus (Keplers SN) [Wikipedia]

$$J = \frac{2}{5} M R_1^2 \cdot \omega_1$$

the angular momentum must be preserved. When the radius R decreases during the collaps to $R_2 \ll R_1$ the angular velocity ω must increase.

The rotational energy

$$E_{\text{rot}} = \frac{1}{2} I \omega^2 = \frac{1}{5} M R^2 \omega^2 = \frac{5 J^2}{4 M R^2} \tag{5.90}$$

increases with $1/R^2$ for a constant angular momentum J.

Example

We assume a star with mass $M = 3M_\odot = 6 \cdot 10^{30}\text{kg}, R_1 = 3 \cdot 10^8\text{m}$ (radius of the central region) and $\omega_1 = 3 \cdot 10^{-6}\text{s}^{-1}$ (as our sun). During the collapse to a neutron star with $R_2 = 10^4\text{m}$ the angular velocity becomes.

$$\omega_2 = \omega_1 \cdot \left(R_1 / R_2\right)^2 = 3 \cdot 10^3\,\text{s}^{-1} \Rightarrow T = 2\,\text{ms}.$$

$$\text{E}_{\text{rot}_1} = \left(\frac{6}{5}\right) \cdot 10^{30} \cdot \left(30 \cdot 10^8 \cdot 3 \cdot 10^{-6}\right)^2 \approx 10^{36}\text{J};$$

$$\text{E}_{\text{rot}_2} = \left(\frac{6}{5}\right) \cdot 10^{30}.$$ ◄

Part of the angular momentum will be carried away by the expelled star envelope during the supernova-explosion. The neutron star therefore has a somewhat smaller rotational energy but still much more than the original star.

If we assume, that 70% of the angular momentum is carried away, the rotational energy is still $\text{E}_{\text{rot}_2} = 10^{44}\text{J}$. This value is also obtained for a neutron start with $T = 2\text{ms}$, $M = 2M_\odot$ and $R = 10\text{km}$.

Note, that the rotational energy of such a neutron star is equal to the total energy, which our sun ($L_\odot = 3.85 \cdot 10^{26}\text{J/s}$) has radiated during its lifetime ($\tau \approx 10^{10}\text{years}$) !!

Now the question arises, under which conditions such a fast rotating star can be stable, although the centrifugal forces are enormous and might pull apart the star.

The centrifugal acceleration at the equator is

$$a_Z = \omega^2 \cdot R, \tag{5.91}$$

while the gravitational acceleration is

$$a_G = -G \cdot \frac{M}{R^2}. \tag{5.92}$$

Replacing the mass M by the density ϱ the stability condition becomes

$$a_G + a_Z \leq 0 \Rightarrow \varrho > \frac{3\omega^2}{4\pi G}. \tag{5.93}$$

Inserting numerical values for $\omega = 3 \cdot 10^3\text{s}^{-1}$ one obtains $\varrho > 3 \cdot 10^{16}\text{kg/m}^3$.

Such high densities are only realized in Neutron Stars (see Sect. 5.7.4).

This implies that pulsars must be rotating neutron stars.

How are the radio pulses generated??

During the collapse of the central region of a massive star the contraction of the plasma (ions and electrons) causes the initially weak magnetic field to be carried along

with the contracting plasma. The magnetic field is therefore also compressed. For a constant magnetic flux the flux density (= magnetic field strength B) increases during the compression. When the star with radius R_1 collapses to a neutron star with radius $R_2 \ll R_1$ the magnetic field strength B increases by the factor $(R_1/R_2)^2$.

Example

Initial magnetic field $B_1 = 10^{-2}$ Tesla, $R_1 = 10^5$ km, $R_2 = 10$ km $\Rightarrow B_2 = 10^6$ Tesla. ◄

▶ **Neutron Stars have a very strong magnetic field.**

Since the initial magnetic field before the collapse generally shows strong local variations (see Sect. 5.4.x), the field after the collapse might not have the form of a magnetic dipole field and its central direction will not necessarily coincide with the rotational axis of the neutron star (Fig. 5.52). The stab points of the magnetic field fixed at a position on the surface of the neutron star then rotate with the star around the rotational axis. The rotating magnetic field can accelerate charged particles, in particular electrons, up to very high energies.

Note: The velocity of a point at the equator of a rotating neutron star is $v = R \cdot \omega \approx 10^4 \cdot 3 \cdot 10^3 \mathrm{m/s} = 3 \cdot 10^7 \mathrm{m/s} = 0.1c$.

According to the Lorentz-force the electron is forced on a helical path around the magnetic field lines and rotates with them around the rotational axis. Its velocity increases with increasing distance from the center of the star and can reach velocities close to c. These relativistic electrons emit synchrotron radiation in a direction, which is limited (as in the synchrotron) within a narrow cone around the velocity of the electrons. Furthermore heavy charged particles can fall onto the surface of the neutron star, due to the strong gravitational force and emit during the impact intense radiation.

The rotating neutron star therefore emits radiation into a narrow cone which rotates with the neutron star, similar to the radiation beam of a rotating dome of a light house. For an observer who just stands at the right place in the plane of the rotating light beam the received radiation appears as a short pulse. Its time duration $\Delta\tau = \Omega/\omega$ depends on the solid angle Ω of the radiation cone and on the angular velocity ω of the rotating neutron star.

Example

$\Omega = 0.1\mathrm{rad}, \omega = 3 \cdot 10^3 \mathrm{s}^{-1} \Rightarrow \Delta\tau = 5 \cdot 10^{-5}\mathrm{s}$. ◄

There is another hint to the assumption that pulsars are rotating neutron stars: We had discussed, that neutron stars are formed together with a supernova explosion, where the envelope of the star is ejected and the remnants can be observed as irregularly shaped nebulas. It immediately suggests itself to look for pulsars in the center of such nebulas. The most famous example is the crab nebula (Fig. 5.50) as remnant of a supernova, observed in 1054. Indeed the first pulsar with a period of $T = 33.2$ms was detected there in 1968.

Meanwhile many hundreds of pulsars have been discovered (Fig. 5.53), although this might be only a small fraction of all existing pulsars, because we can see them only, if their emission cone hits the earth.

5.7.5 Black Holes

If the gravitational pressure of a star with mass M exceeds the Fermi-pressure of the neutrons, there is no further force that might compensate the gravitational pressure and the star collapses to a black hole.

We will now discuss in more detail the question: What are the characteristic features of a black hole and how can it be formed?

5.7.5.1 The Collapse into a Black Hole

Although the processes that lead to a black hole have to be discussed, strictly speaking, within the framework of the General Relativity, a nonrelativistic calculation brings already realistic results, which are confirmed in the essential points by a strict calculation.

An atom with mass m, sitting at the surface of a spherical mass

$$M = (\frac{4}{3})\pi\varrho R^3$$

experiences the gravitational force

$$F_G = G \cdot m \cdot \frac{M}{R^2}$$

And therefore obey the equation of motion

$$m \cdot \frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = -G \cdot \frac{M}{R^2} \cdot m. \tag{5.94}$$

Multiplication with $\mathrm{d}R/\mathrm{d}t$ and division by m yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 - \frac{M \cdot G}{R}\right] = 0, \tag{5.95}$$

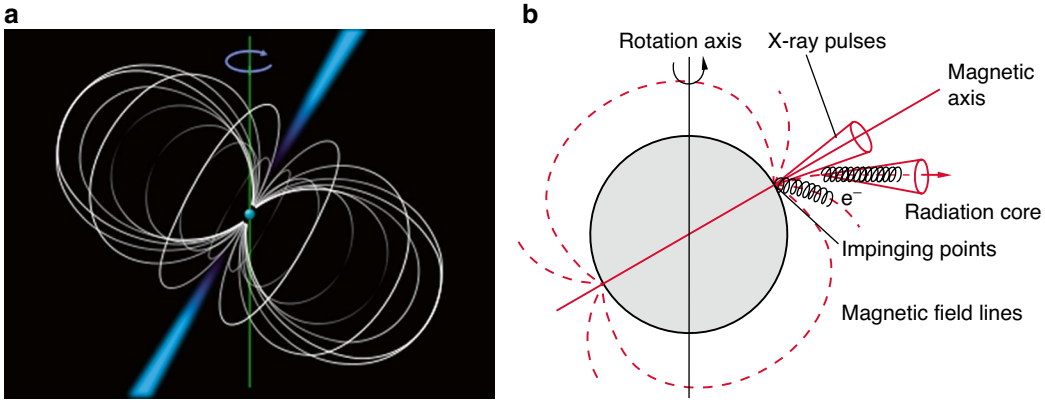


Fig. 5.52 **a** Schematic illustration of the emitted radiation cone from a rotating neutron star (Made by Mysid in Inkscape schematic.jpg by Roy Smits.wikimedia). **b** Magnetic field of a rotating neutron star and possible mechanism of the emission of radiation

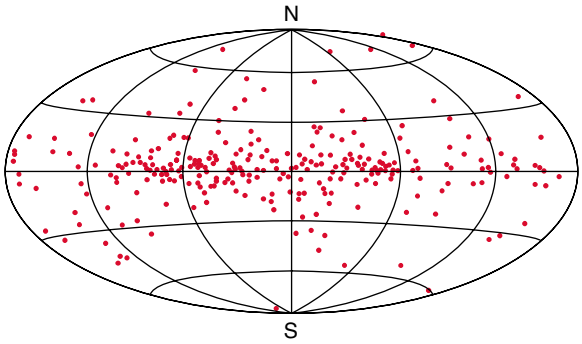


Fig. 5.53 Spatial distribution of the 300 pulsars found up to date. They are concentrated along the central plane of our milky way (after R. Kippenhahn, A Weigert: Stellar Structure and Evolution. Springer Heidelberg 1984)

which gives the energy conservation law

$$\frac{1}{2}\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 - \frac{M\cdot G}{R} = \text{const.} = -\frac{M\cdot G}{R_0} \qquad (5.96)$$

where $(t_0) = 0$.

Integration of (5.96) gives

$$t(R) = \left(\frac{R\cdot R_0(R_0\cdot R)}{2M\cdot G}\right)^{1/2} + R_0\left(\frac{R_0}{8MG}\right)^{1/2}\arccos\left(\frac{2R}{R_0}-1\right), \qquad (5.97)$$

If the collaps starts at $R = R_0$ it is $t(R_0) = 0$. For $R = 0$ we get the time t_c the object takes to collapse from the edge of the star core with $R = R_0$ to the singularity at $R = 0$.

$$t_c = \pi\left(\frac{R_0^3}{8G\cdot M}\right)^{1/2} \approx \frac{1}{\sqrt{G\cdot\varrho}}. \qquad (5.98)$$

Note: The same result has been obtained in Sect. 5.4 for the contraction of a gas cloud to form a protostar. It is remarkable that the collapse time does not depend on the radius of the contracting cloud or star but only on the density ϱ . There is a great difference between the two scenarios: At the formation of a protostar the density ϱ is smaller by many orders of magnitude than for the collapse to a black hole. Therefore the collapse times are much longer and the velocities of the collapsing particles $v \ll c$ are much smaller than the velocity c of light while in the case of the black hole $v \to c$ as soon as the radius R becomes smaller than the Schwarzschild radius $\mathcal{R}_S$ (see next section).

Example

Assume that the nuclear fusion processes in the central part of a star with $M = 10M_\odot$ and mean density $\varrho = 10^4\mathrm{kg/m}^3$ suddenly stop. Now the gravitational pressure cannot be any longer compensated and the initially stable star starts to collapse. The collapse time then is, according to (5.98).

$$t_c = \frac{1}{\sqrt{G\cdot\varrho}} = 1200\ s = 20\ \text{min}.$$

Actually only the central part of the star collapses. When we take a radius $R = 10^3$km and a mass $M = 2\,M_\odot$ the collapse time becomes $t_c = 70$ms. Before the star envelope with its large collapse time can contract, the supernova-explosion blows the envelope away. ◄

For $M = 10\,M_\odot$ indeed such a collapse of the core occurs after the nuclear fusion processes end. The classical gravitation theory predicts that the whole central mass collapses to a point with radius $R = 0$, which means that the density becomes infinite ($\varrho = \infty$). This classical description

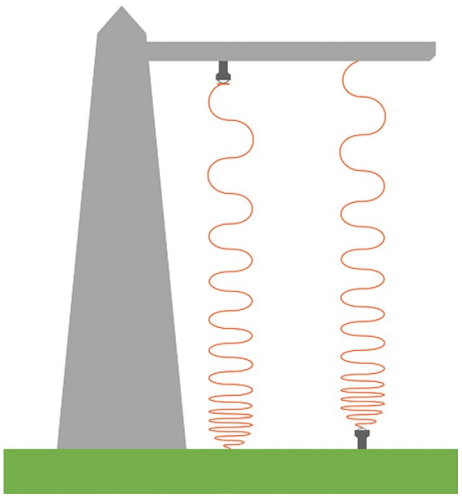


Fig. 5.54 Schematic illustration of the frequency shift of photons rising in a gravitational field (Wikipedia)

neglects two aspects: Firstly the velocity of the collapsing particles approaches the velocity of light and therefore one has to treat the problem by a relativistic theory. Secondly the concentration of a mass to a mass singularity demands the treatment with the general relativity because the mass concentration deforms the metric of space. However, also the general relativity theory predicts a mass singularity. But describes the formation of it in a different way.

Whether there would also be a mass singularity within the framework of a relativistic quantum theory, is not fully known up to date. In order to better understand the course of the collapse into a black hole we will at first introduce the Schwarzschild radius.

5.7.5.2 Schwarzschild Radius

In Volume 3, Sect. 3.1.5 it has been explained that the energy $E = h \cdot \nu$ of a photon decreases by the amount

$$\Delta E = \frac{h \cdot \nu}{c^2}(\phi_G(r_2) - \phi_G(r_1)) \tag{5.99}$$

when it rises in a gravitational potential $\phi(R)$ from the radius r_1 to the larger radius r_2 (red shift in the gravitational potential) (Fig. 5.54).

The potential energy of a mass m in the gravitational potential ϕ is $E_{pot} = m \cdot \phi$.

The photon mass is $m_{ph} = h \cdot \nu / c^2$. Its mass change is then

$$\Delta m = \left(\frac{m}{c^2}\right) \cdot \Delta\phi \rightarrow \Delta E = \Delta m \cdot c^2 = \left(\frac{h \cdot \nu}{c^2}\right) \cdot \Delta\phi$$

With $\Delta E = h \cdot \Delta\nu$ this corresponds to a relative frequency shift

$$\frac{\Delta\nu}{\nu} = \frac{\Delta E}{E} = \frac{\Delta E}{h \cdot \nu} \text{ with } \Delta\nu = \nu_0 - \nu \tag{5.100}$$

$$\rightarrow \frac{\Delta\nu}{\nu} = \frac{\Delta\phi}{c^2}.$$

Example

For the heights H = 20 m the potential difference in the gravitational field of the earth is $\Delta\phi = g \cdot H = 9.81 \cdot 20\ m^2\ s^{-2}$. The frequency shift is then $196.2/(9 \cdot 10^{16}) = 2.2 \cdot 10^{15} \cdot \nu$. For $h \cdot \nu = 10$ keV (γ-radiation) is $\nu = 5 \cdot 10^{18} \rightarrow \Delta\nu \approx 10^4\ s^{-1}$. Such small relative frequency shifts $\delta/\nu = 2 \cdot 10^{-15}$. can be only measured with special techniques e.g. the Mößbauer Spectroscopy. ◄

This relation (5.100) was confirmed by the experiment performed by *Pound* and *Rebka,* who send **γ-**radiation from the ground on earth ($r = 0$) into a detector at the heights H = 22.6 m [15]. The source was a radioactive Iridium sample and the detector was a Mößbauer arrangement which could measure the frequency shift by the Doppler- shift of the absorption line Nowadays (Fig. 5.55).

The gravitational potential at the surface of a spherical mass M and radius R is

$$\phi(R) = -\frac{G \cdot M}{R}. \tag{5.101}$$

When a photon rises from the surface to $R = \infty$ is $\Delta\phi = -G \cdot M/R$ and the redshift of the photon becomes

$$\frac{\Delta\nu}{\nu} = -\frac{G \cdot M}{Rc^2}. \tag{5.102}$$

When we define the **Schwarzschild-radius**

$$\mathcal{R}_S = \frac{2MG}{c^2} \tag{5.103}$$

we get the frequency shift

$$\frac{\Delta\nu}{\nu} = -\frac{\mathcal{R}_S}{2R}. \tag{5.104}$$

Light which is emitted from the Schwarzschildradius $\mathcal{R}_S$ has a frequency shift $\Delta\nu = -\frac{1}{2}\nu$. Its wavelength λ increases to 2λ.

While (5.104) gives the correct result for $R \gg \mathcal{R}_S$, for ranges $R \approx \mathcal{R}_S$ one has to use relativistic calculations because the space curvature is no longer negligible in these ranges. The relativistic calculations give instead of (5.104)

$$\frac{\Delta\nu}{\nu} = -1 + \sqrt{1 - \mathcal{R}_S/R} \tag{5.105}$$

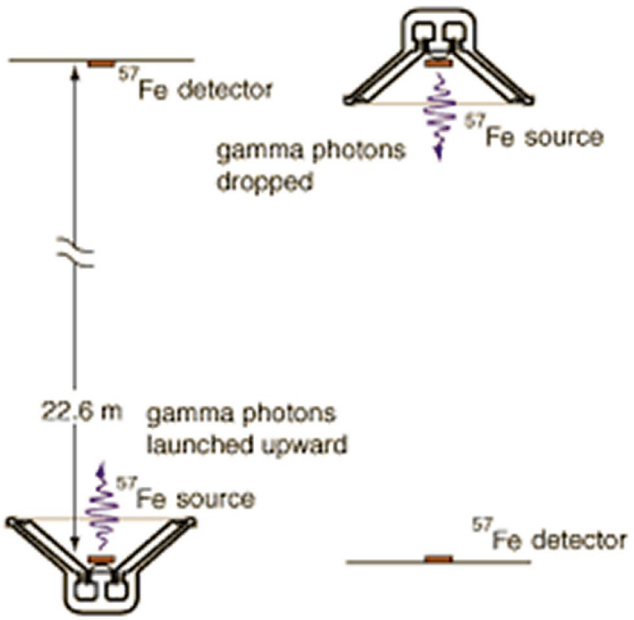


Fig. 5.55 Experimental setup by Pound and Rebka for measuring the frequency shift of photons uprising in the earth gravitational field (Wikipedia)

For $R = \mathcal{R}_S$ the frequency shift becomes $\Delta\nu = -\nu$ which means that $\nu \to 0$ for $R \to \mathcal{R}_S$.

For $R < \mathcal{R}_S$ the radicand in (5.105) becomes negative and the frequency ν becomes imaginary, which implies that no radiation will escape the volume inside the Schwarzschild radius $\mathcal{R}_S$.

For $R \gg \mathcal{R}_S$ we can expand the square root

$$\sqrt{1 - R/\mathcal{R}_S} \approx 1 - \frac{1}{2}R/\mathcal{R}_S$$

and the relativistic calculation merge into the classical result (5.104).

In order to escape from the volume inside the Schwarzschild-radius the kinetic energy of a particle with mass m must exceed the negative potential energy

$$\frac{1}{2}m \cdot v^2 > -E_{pot} = \frac{G \cdot M \cdot m}{R} \to \mathrm{v} > \sqrt{2G \cdot M/\mathrm{R}}. \tag{5.106}$$

For $R = \mathcal{R}_S \to \mathrm{v} > \mathrm{c}$.

which means that particles can only escape, if their velocity is larger than the velocity c of light. Since this is impossible we can conclude:

- Particles and also photons within the Schwarzschild-radius $\mathcal{R}_S$ cannot escape from the volume inside $\mathcal{R}_S$, because they would need a velocity v larger than the velocity of light c. Therefore this volume is called a **black hole**.

Inserting into (5.103) the numerical values one obtains for the Schwarzschild-radius

$$\mathcal{R}_S = M \cdot 1.48 \cdot 10^{-27}\mathrm{m},$$

where M is given in kg.

The Schwarzschild volume is then

$$V_S = \left(\frac{4}{3}\right)\pi\mathcal{R}_S^3$$

And the density

$$\varrho_S = \frac{M}{V_S} = \frac{3c^6}{(32\pi G^3 M^2)}. \tag{5.107}$$

- As soon as the density of the core of a collapsing star with mass M exceeds the critical density ϱ_S, it collapses into a black hole.

In Table 5.8 the numerical values of the Schwarzschild radius are given for some cosmic objects.

Example

With the sun mass $M_\odot = 1.989 \cdot 10^{30}$ kg we get $\mathcal{R}_S = 2.95 \cdot (M/M_\odot)$ km. The Schwarzschild radius of our sun is then less than 3 km! The sun must be compressed into a volume with radius less than 3 km in order to become a black hole. The Schwarzschild radius of the earth ($M = 5.9 \cdot 10^{24}$ kg) is only about 9 mm! ◄

The minimum density of a black hole with mass M is

$$\varrho > \varrho_S = \frac{M}{\frac{4}{3}\pi\mathcal{R}_S^3} = \frac{3MM_\odot^3}{4\pi \cdot 2.95^3 \cdot M^3} = 1.84 \cdot 10^{16}(M_\odot/M)^2 \mathrm{g/cm}^3. \tag{5.108}$$

For very massive black holes the density can be smaller than that of a white dwarf.

The Schwarzschild radius defines the surface of a non-rotating black hole. It is called the **event horizon**.

For rotating black holes the mass cannot collapse to a singularity with $R = 0$, because the angular momentum has to be preserved during the collapse and for $R = 0$ it would become infinite.

5.7.5.3 Light Deflection in a Gravitation Field

When a light beam passes through the gravitational field of a mass M it will be bent by the gravitation (Fig. 5.56).

We can calculate this deflection approximately as follows [16].

Table. 5.8 Schwarzschild-Radius $\mathcal{R}_S$ of some celestial objects

Object	Mass m (kg)	Radius R (m)	$\mathcal{R}_S$(m)	$\mathcal{R}_S/R$
Earth	$6 \cdot 10^{24}$	$6 \cdot 10^{6}$	$9 \cdot 10^{-3}$	$1.5 \cdot 10^{-9}$
Sun	$2 \cdot 10^{30}$	$7 \cdot 10^{8}$	$3 \cdot 10^{3}$	$4 \cdot 10^{-6}$
White dwarf	$2 \cdot 10^{30}$	10^{7}	$3 \cdot 10^{3}$	$3 \cdot 10^{-4}$
Neutron star	$3 \cdot 10^{30}$	10^{4}	$4.5 \cdot 10^{3}$	0.45
Galaxy	pt10^{41}	10^{21}	10^{14}	$10-^{7}$

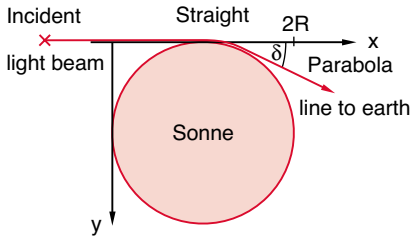


Fig. 5.56 Deflection of a light beam on its path closely along the sun

Since photons can be treated in gravitational fields like particles with the mass $m = h\nu/c^2$ (see Sect. 5.7.5.1) the acceleration in y-direction $g = G \cdot M/R^2$ acts on a photon passing into the x-direction close to a massive object, e.g. our sun. The trajectory of the photon in the x–y-plane is then

$$x = c \cdot t; y = \frac{g}{2}t^2 = \frac{G \cdot M}{2R^2}t^2. \qquad (5.109)$$

Since the velocity in x-direction is very large compared with that in y-direction, the velocity v_x will not be much affected by the deflection and we can set $v_x = c$. Inserting $t = x/c$ in the equation for y in (5.109) we get the parabola

$$y = \frac{G \cdot M}{2R^2c^2}x^2. \qquad (5.110)$$

Most of the deflection will occur in the range 0<x<2R. The deflection angle is then approximately

$$\tan\delta \approx \delta = y\prime(2R) = \frac{2MG}{Rc^2}. \qquad (5.111)$$

This result was obtained by a classical calculation which takes only into account the deflection during the path-length $\Delta x = 2R$ in order to get the approximate value of the deflection.

The exact calculation within the framework of the general relativity theory gives a deflection which is twice as large as the value (5.111).

With the Schwarzschild radius $\mathcal{R}_S$ one obtains for $\delta \ll \pi$ i.e. $R \gg \mathcal{R}_S$ the correct deflection angle.

$$\delta = \frac{4GM}{Rc^2} = \frac{2\mathcal{R}_S}{R}. \qquad (5.112)$$

▶ The light deflection in the gravitational field of a mass M is equal to twice the ratio of Schwarzschild radius to the distance R from the mass center.

This theoretical result was experimentally confirmed when during a sun eclipse the light from a star was observed which passes on its way to the earth close to the edge of the sun (Fig. 5.56). Initially the optical measurements had been not very accurate. However, recently the accuracy could be considerably improved and yielded the deflection angle $\delta = 1.''75$.

The most accurate results are obtained with radio-astronomy, where the radio signals from a far distant object (e.g. a quasar = Quasi-Stellar Radio source) are measured directly before and after their occultation by the sun.

For the very accurate determination of star positions by the astrometric Satellites HIPPARCOS and GAIA the light deflection by the sun and even by planets r, such as Jupiter, or other large object has to be taken into account, even for large angular distances [17].

5.7.5.4 Time Progression During the Collapse into a Black Hole

The time progression of the collapse when the central region of a massive star collapses into a black hole after the end of the nuclear fusion processes is illustrated in Fig. 5.57 for an observer 2 far away from the Schwarzschild radius and for an observer 1 falling with the collapsing mass into the black hole. The vertical central line in Fig. 5.57 is the world line of the star center (see Vol. 1, Sect. 3.6.2), while the circles around the center represent the Schwarzschild radius $\mathfrak{R}_S$(t). The time dependent expression $R(t)$ give the distance from the center $R = 0$ for the different times t during the collapse. The curved line at the distances $R(t)$ from the central line give the path of a hypothetical observer 1 at the edge of the black hole who collapses with the black hole towards $R(t) = 0$. After the collapse time

$$t_c = (G \cdot \varrho)^{-1/2}$$

the star has collapsed to a singularity $R = 0$ and the observer 1 has experienced the same fate.

The second observer 2 sits far outside at $r \gg \mathcal{R}_S$ and therefore does not participate in the collapse. His world line is the vertical straight line r = const. at the right side of the figure.

Now observer 1 has arranged with observer 2 that he sends signals with regular time intervals Δt_1 to observer 2. He sends then during his collapse time altogether $N = t_c/\Delta t_1$ signals. However, these signal do not reach observer 2 in equal time intervals, because the light path through gravitational fields is curved and the red world lines

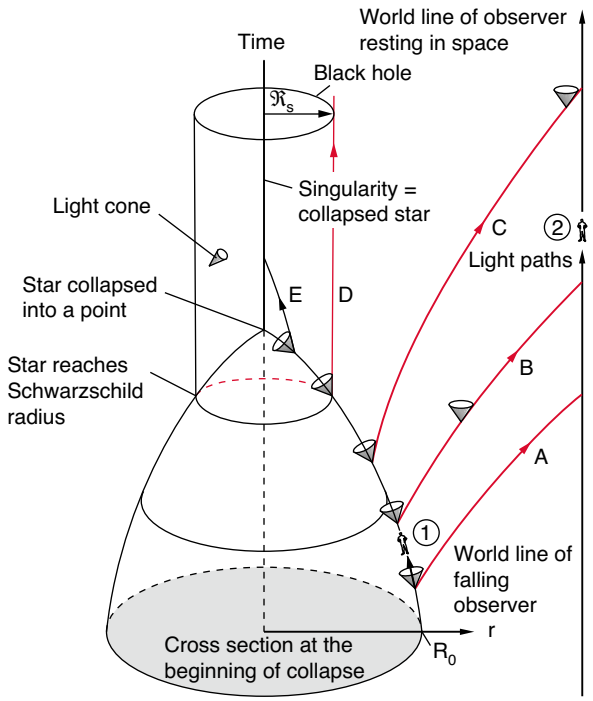


Fig. 5.57 Time progression of the collapse of a black hole shown in a x-t-diagram (Minkowski diagram) for two different observers (after R. and H. Sexl, Weiße Zwerge-Schwarze Löcher. Vieweg Braunschweig 1995)

of the signals in Fig. 5.57 show a different slope depending on the distance of observer 1 from $R = 0$.When observer 1 reaches the Schwarzschild radius the world line of the signals become a vertical straight line because the light cannot leave the volume with $R \leq \mathcal{R}_S$. The time intervals between the signals received by 2 are becoming longer and longer until 2 receives no signal any longer as soon as observer 1 has arrived at $R = \mathcal{R}_S$, when the light path becomes a circle with radius $R = \mathcal{R}_S$ around the central mass. Observer 2 therefore can trace the collapse only until observer 1 reaches $R = \mathcal{R}_S$ but not until $R = 0$.

For the external observer 2 the collapse slows down until it comes to a halt at $R = \mathcal{R}_S$ (Fig. 5.58).

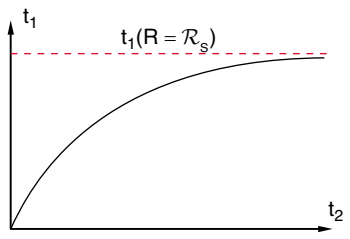


Fig. 5.58 Relation between the times measured by the observers A and B

A more detailed calculation shows that a signal that is sent from 1 at the radius $R = \mathcal{R}_S(1 + \varepsilon)$ with $\varepsilon \ll 1$ reaches observer 2 only after a time $= -(R/c) \cdot \ln\varepsilon$.

For the observer 2 the collapsing star never reaches the Schwarzschild radius, which implies that the star never becomes a black hole. The signals sent by observer 1 are more and more redshifted, since 1 moves with increasing velocity away from 2. This means that the star becomes more and more dark and for $\lambda \to \infty$ the luminosity tends to zero. Its time-dependent luminosity is

$$L(t) = L_0 \cdot e^{-t/\tau} \text{ with } \tau = \mathcal{R}_S/c \qquad (5.113)$$

where $\tau = t_c$ (5.98) is the collaps time and the time zero is $t = 0$ for $R = R_0$.

Example

For $M = 10M_\odot$ is $\mathcal{R}_S = 30\text{km} \Rightarrow \tau = 10^{-4}\text{s}$. ◄

5.7.5.5 The Search for Black Holes

Since one cannot see black holes, as has been explained above, one has to search for indirect evidence of their existence. There are several possibilities:

If the black hole is one of the components of a binary system and the other component is a visible star, both components orbit around their common center of mass. The observable proper motion of the visible component allows one to determine the mass of the black hole (see Sect. 5.1). Such a situation is indeed realistic. Even if both components have been formed at the same time, the more massive component develops faster and comes earlier to its end, forming a black hole if its mass exceeds the Oppenheimer-Volkow limit while the less massive component might be still in the main sequence stage if it has not been teared apart by the supernova explosion of the other component.

Another possibility of detecting black holes are binaries with very close components, where the more massive component has already developed into a black hole. If the distance between the components is smaller than the Roche-limit (see Sect. 5.5.x), mass is flowing from the lighter to the more massive component.

During this flow the particles are accelerated to relativistic velocities and emit radiation in the X-ray spectral range which can be detected, as long as the particles are still outside the Schwarzschild radius. Such binaries are called X-ray binaries. One example is the X-ray source in the Cygnus constellation about 8000 ly apart which could be identified as an X-ray binary with a black hole as one of the components.

Fig. 5.59 The supermassive black hole at the core of the supergiant elliptical galaxy Messier 87, with a mass of about 7 billion times that of the Sun, as depicted in the first false-colour image of radio waves released by the Event Horizon Telescope (10 April 2019). Visible are the crescent-shaped emission ring and a central shadow, which are gravitationally magnified views of the black hole 's photon ring and the photon capture zone of its event horizon. The crescent shape arises from the black hole 's rotation and relativistic beaming; the shadow is about 2.6 times the diameter of the event horizon (By Event Horizon Telescope, https://www.eso.org/public/images/eso1907a/ (image link))

A very massive black hole has been detected in the center of our galaxy which has a mass of $M = 3.7 \cdot 10^6 M_\odot$ (see Sect. 6.5) and there is evidence that many galaxies have a black hole in their center.

A very convincing proof of a black hole has been recently given by the detection of gravitational waves emitted during the merging of a black hole with a neutron star [5.18].

A photo of a massive black hole is shown in Fig. 5.59.

5.8 Quasars

Quasars are radiation sources with a very high luminosity but a small spatial extension. They had been first discovered in the radio frequency range (quasars = quasi-stellar radio sources), but meanwhile it turned out that they emit radiation in all spectral ranges, also in the visible. Because of their small spatial extension it was at first assumed that the sources of this very intense radiation were special stars. Detailed investigations have, however, confirmed, that the origin of their radiation are black holes in the center of galaxies. They suck in particles which are accelerated to relativistic velocities and emit intense radiation. The spectra show broad emission lines, which indicates that the emitting particles must have a broad random velocity distribution. The particles falling into the black hole come from regions of the rotating galaxy outside of the black hole. Because their angular momentum must be preserved they cannot fall radially into the black hole but must move on helical paths with decreasing radii into the black hole. This leads to a large acceleration and the emission of bremsstrahlung which extends from the radio frequency range to the X-ray and even gamma region.

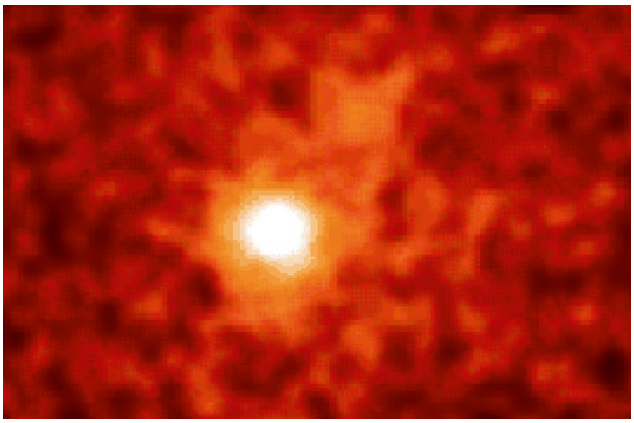


Fig. 5.60 Quasar 3C79 which is 5 billion ly away (Compton Observatory NASA)

In 2010 it was discovered that a quasar at the distance of 1.6 billion light years acted as gravitational lens for the radiation from a galaxy at a much larger distance of $5.9 \cdot 10^9$ ly. This allowed the determination of the quasar mass [19].

The luminosity of quasars is about 10^{12}–10^{14} times larger than that of our sun. Therefore it is possible to detect quasars even at very large distances up to $13 \cdot 10^9$ ly. The radiation of such quasars had been emitted 13 billion years ago. The investigation of their radiation therefore allows a glimpse into the early universe not long after its creation. Figure 5.60 shows a quasar detected with the Compton Gamma Telescope, on a satellite, which has orbited the earth from 1991–2000 and measured the γ-radiation which is not transmitted by our atmosphere and cannot therefore be detected by earth-bound telescopes.

5.9 Observable Phenomena During the Final Stage of Stars

We have discussed in the last sections that stars after their stay on the main sequence pass through different instable phases depending on their mass. These phases become noticeable by fluctuations of their luminosity. Stars in such phases are called **variable stars**. Sometimes these variations of the luminosity are dramatic when it varies by many orders of magnitude. Examples are novae and supernovae.

Up to now more than 100.000 variable stars have been observed and the analysis of the data from GAIA will certainly increase this number considerably. In this section we will shortly discuss the physical causes for such changes of the luminosity. The most accurate measurements are nowadays performed with CCD-detectors and image intensifiers.

The observed time dependent apparent luminosity of a star is called its **light curve**. For periodic changes the amplitude and the period of the luminosity changes $\Delta L(t)$ can be determined.

5.9.1 Pulsating Variable Stars

The pulsating variable stars represent the by far most abundant group of variable stars (about 90%). For these stars one observes a periodic variation of the luminosity and also of the wavelength of emitted spectral lines. This indicates that the radius of the star experiences periodic changes, i.e. the star represents and radially oscillating sphere (Fig. 5.61). The changing surface area results in corresponding changes of the luminosity and the pulsating star radius causes periodic Doppler-shifts of the spectral lines. The observed Doppler-shifts corresponds to velocities of the pulsating star surface between 40 and 200 km/s.

The group of pulsating variable stars includes the *Cepheides* (named after the star δ-cephei in the constellation *Cepheus*) where the pulsation was first discovered; the Mira-Variables (named after the star Mira Ceti in the constellation Cetus) and the RR-Lyra stars (in the constellation Lyra, which are often found in the globular star clusters surrounding our galaxy.

The Cepheides have pulsation periods between 2 and 50 days. The periodic light curve shows a steep rise and a more flat decline (Fig. 5.61). The integration of the radial velocities $v_r = dR/dt$ over ¼ of the period gives the amplitude ΔR of the pulsation R(t), which can amount up to $0.1R$. This means that the volume of the star changes by 30% and the surface by 20%! The maximum of the luminosity appears at the time of the maximum expansion-velocity not at the maximum surface as one would expect! The reason for this unexpected result is that the maximum temperature of the star surface occurs at the time of the maximum expansion velocity. Since the luminosity L is proportional to $T^4 \cdot R^2$ the temperature is more important for L than the surface area.

The star oscillation is fuelled by absorption of the radiation from the central part by neutral and ionized hydrogen $H + H^+$ and helium atoms $He + He^{++}$ in a layer closely below the star surface. The resulting heating leads to an expansion of the star envelope. This in turn results in a decrease of the temperature and gas density and therefore

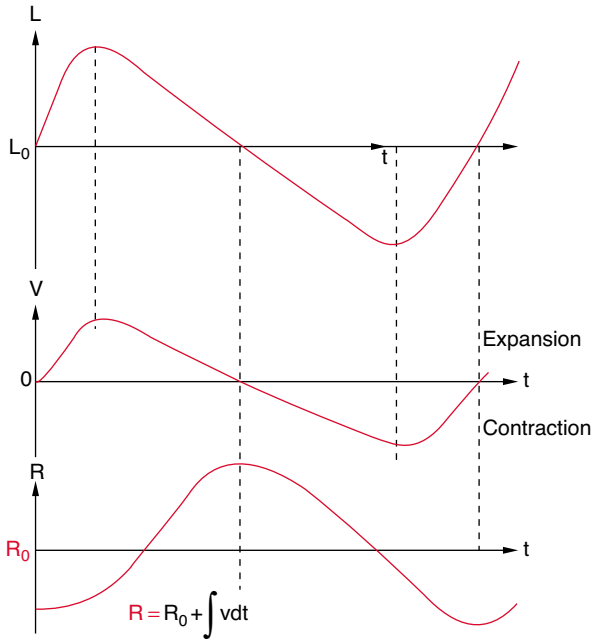


Fig. 5.61 Luminosity L(t), radial velocity $v_r(t)$ and star radius R(t) of a periodically pulsating star

also a decrease of the absorption. The envelope starts to contract again, until the processes of absorption and expansion start again.

The pulsation period corresponds to the period of a radial acoustic resonance oscillation of the star. We will give a short derivation [6].

We regard a volume element with the mass $dm = \varrho dV$ at the distance r from the center of the star. Its equation of motion under the influence of gravitational and pressure force is

$$\varrho \frac{\partial^2 r}{\partial t^2} = -G \cdot \frac{M \cdot \varrho}{r^2} - \frac{\partial p}{\partial r}. \tag{5.114}$$

For the equilibrium condition (r_0, p_0, ϱ_0) the volume element dV should be balanced, i.e. the total force is zero. Now we regard a small perturbation which changes the equilibrium state (r_0, p_0, ϱ_0) into the state (r, p, ϱ) with

$$\left.\begin{aligned} r &= r_0[1 + \xi(t, r_0)] \\ \varrho &= \varrho_0[1 + \eta(t, r_0)] \\ p &= p_0[1 + \zeta(t, r_0)] \end{aligned}\right\} \text{with } \xi, \eta, \zeta \ll 1. \tag{5.115}$$

The displacement from the equilibrium leads to restoring forces which induce the oscillation. Inserting (5.115) into (5.114) we get, when taking into account the constant mass of the star

$$\begin{aligned} M(r) &= M(r_0) \\ \Rightarrow dM(r) &= 4\pi r^2 \varrho \; dr = 4\pi r_0^2 \varrho_0 \; dr_0 \\ \Rightarrow \partial r/\partial r_0 &= \tfrac{r_0^2 \varrho_0}{r^2 \varrho} = \tfrac{1}{(1+\xi)^2(1+\eta)} \approx \tfrac{1}{1+2\xi+\eta} \end{aligned} \tag{5.116}$$

and according to Eq. (4.56)

$$\frac{\partial p}{\partial r} = -GM\varrho/r^2 \Rightarrow \frac{\partial p_0}{\partial r_0} = -GM\varrho_0/r_0^2 \quad (5.117)$$

The three terms in the equation of motion (5.114) then become

$$\begin{aligned}\frac{\partial p}{\partial r} &= \frac{\partial p}{\partial r_0}\cdot\frac{\partial r_0}{\partial r} \\ &= \left[\frac{\partial p_0}{\partial r_0}(1+\zeta) + p_0\frac{\partial \zeta}{\partial r_0}\right]\cdot(1+2\xi+\eta) \quad (5.118a)\\ &= \left[\frac{-GM\varrho_0}{r_0^2}(1+\zeta) + p_0\frac{\partial \zeta}{\partial r_0}\right]\cdot(1+2\xi+\eta),\end{aligned}$$

$$G\frac{M}{r^2}\varrho = G\frac{M}{r_0^2}\varrho_0(1-2\xi+\eta), \quad (5.118b)$$

$$\varrho\frac{\partial^2 r}{\partial t^2} = \varrho_0 r_0\frac{\partial^2 \xi}{\partial t^2}. \quad (5.118c)$$

We then get the equation of motion

$$\frac{\partial^2 \xi}{\partial t^2} + G\frac{M}{r_0^3}(4\xi+\zeta) + \frac{p_0}{\varrho_0 r_0}\frac{\partial \xi}{\partial r_0} = 0. \quad (5.119)$$

When we regard a mass element $dm = \varrho dV$ closely below the surface of the star where temperature T and pressure p are much smaller than in the central part, we can apply the classical gas equation

$$p_0 = (\varrho_0/(n\cdot m_p))\cdot kT_0$$

which tells us that because of the small value of T_0 in the outer envelope, the ratio p_0/ϱ_0 must be much smaller than in the central part. This means that we can neglect in (5.119) the last term and we get for $r_0 = R$ the relation

$$\frac{\partial^2 \xi}{\partial t^2} = -G\frac{M}{R^3}(4\xi+\zeta). \quad (5.120)$$

If the displacement from the equilibrium state occurs so fast that no energy is exchanged with the surrounding the change is adiabatic. With the adiabatic equation

$$(p/p_0) = (\varrho/\varrho_0)^\kappa \text{ with } \kappa = c_p/c_V \quad (5.121)$$

We get

$$\begin{aligned}(1+\zeta) = (1+\eta)^\kappa &\approx 1+\kappa\eta \\ \Rightarrow \zeta &\approx \kappa\cdot\eta.\end{aligned} \quad (5.122)$$

Since $\varrho \propto r^{-3}$ it follows from (5.115) $\eta = -3\xi$ and we get with (5.120) the equation of the oscillation

$$\frac{\partial^2 \xi}{\partial t^2} = -G\cdot\frac{M}{R^3}(3\kappa-4)\cdot\xi \quad (5.123)$$

Its solution with the initial condition $\xi(0) = 0$ is

$$\xi = A\cdot\sin\omega t \Rightarrow R = R_0(1+A\sin\omega t) \quad (5.124)$$

With the oscillation frequency

$$\Omega = \sqrt{G\cdot\frac{M}{R^3}(3\kappa-4)} \quad (5.125a)$$

And the oscillation period

$$\begin{aligned}\tau = \frac{2\pi}{\omega} &= 2\pi\cdot\sqrt{\frac{R^3}{G\cdot M\cdot(3\kappa-4)}} \\ &= \sqrt{\frac{3\pi}{3\kappa-4}}\cdot\sqrt{\frac{1}{G\cdot\overline{\varrho}}},\end{aligned} \quad (5.125b)$$

where we have inserted the mass $M = \frac{4}{3}\pi R^3\overline{\varrho}$ of the star.

Example

When we insert the numerical values for a red giant star $M = M_\odot$; $R = 10R_\odot$ and $\kappa = 5/3$ (atomic hydrogen) we get the oscillation period $\tau = 3$d. ◄

Since for a constant average temperature the luminosity

$$L = 4\pi R^2\sigma T^4 \quad (5.11.91)$$

is proportional to R^2 its luminosity oscillates with the oscillation period of the surface. The radius R of the star can change by as much as 30%. The variation of the luminosity is caused by the pulsation of the radius as well as that of the temperature related to the changing radius.

One might at first think, that the oscillations are damped because the temperature rise in the volume element dV during the compression phase is diminished by heat conduction to the surrounding. However, this loss of the temperature increase can be compensated by the larger absorption of the radiation from the central part of the star due to the higher gas density at the lower temperature in the volume element dV. The energy periodically supplied to dV keeps the oscillation running.

The important point is the relation between the period T and the absolute luminosity L (**Period-luminosity relation**). If the mass M of the pulsating star is known (e.g. if the star is a component of a binary system) the observed period T allows the determination of the radius R, and therefore of the surface $4\pi R^2$ and the luminosity $L = 4\pi R^2\sigma T^4$ if the temperature T is known, for instance from spectroscopic measurements.

With the help of reference pulsating stars the period-luminosity relation is calibrated and one obtains for

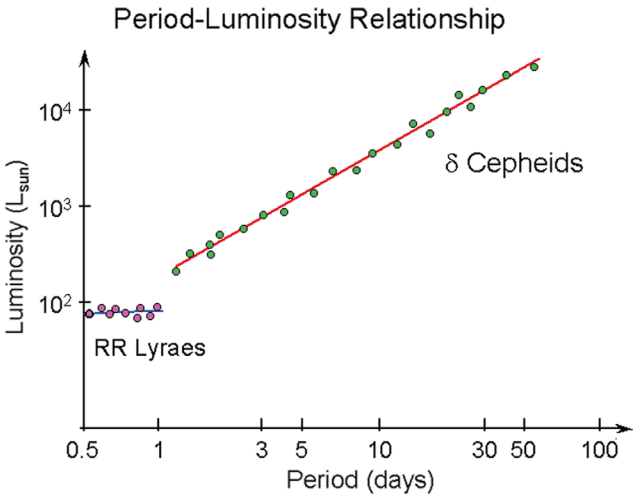


Fig. 5.62 Period-Luminosity relation for δ-Cepheids and RR-lyra Stars (Von R. J. Hall, CC BY-SA 3.0, https://commons.wikimedia.org/)

instance for the Cepheides the relation illustrated in Fig. 5.62.

Measuring the pulsation period of an arbitrary member of the Cepheides, its absolute luminosity can be therefore determined.

The comparison between this known absolute luminosity with the relative luminosity measured on earth, gives the distance of the star according to (3.x).

Periodically pulsating stars are therefore distance meters beyond the distances where trigonometric methods fail. With the Cepheid-method the distance of the nearby galaxies could be reliably measured for the first time.

Note: With the astrometric Satellites HIPARCOS and GAIA the distances of celestial objects up up to 10^6 parsec could be measured with high precision (see Sect. 5.3.y). The Cepheide distance method is therefore no longer up to date.

The **Cepheides** which are supergiants (Fig. 5.16) and belong to the spectral types F–K represent the most important group of pulsating variable stars. Their pulsation periods range from $T = 1$d to $T = 50$d and the amplitudes of the luminosity oscillations are 0.3–0.25 mag (see Sect. 5.3.z).

The **Mira** variable stars are long term variables with periods between 80 and 1000 d (Fig. 5.63). They belong to the giants or supergiants of the spectral type K5–M9. They are therefore colder than the Cepheides.

The colder variable stars are, the longer is their pulsation period.

The amplitudes of the oscillations of the visual radiation are between 2.5 and 11 mag. They are larger than for the Cepheides. The large oscillations of the luminosity are not only due to changes of the surface area but mainly to

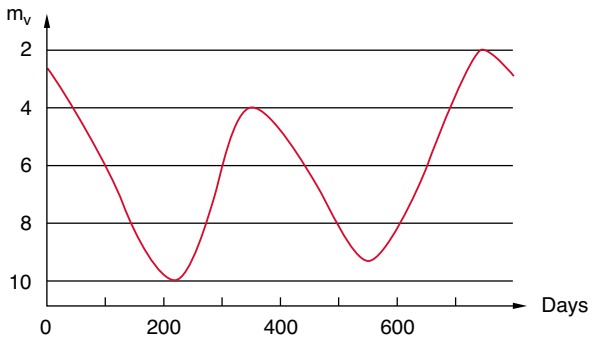


Fig. 5.63 Typical light curve for Mira stars

the larger temperature oscillations. During the cooler phase molecules are formed by recombination of atomic hydrogen H to molecules H_2. This can be confirmed by observations of the molecular hydrogen bands which are much more abundant than in the Cepheides.

A further group of variable stars are the **RR-Lyra** stars with much shorter oscillation periods (T < 1d, Fig. 5.64), which can be found in globular clusters (See Sect. 6.5). They are giant stars of spectral types A–F with small masses $M \sim 0.5\,M_\odot$ which are in a stage shortly after leaving the main sequence, when the Helium-burning has just started. Their radius is about $R = 5R_\odot$.

All pulsating stars represent unstable phases of the star development on their way away from the stable stage on the main sequence to their stable final end stage. In this region of instability in the HRD (Fig. 5.33b) mainly stars are found with unstable radial profiles of pressure and temperature in the outer layers. Small radial perturbations (e.g. pressure waves) may result in large fluctuations of radius R and temperature T and therefore also to fluctuations of the luminosity $L = 4\pi R^2 \sigma T^4$. The resulting change of the absorption coefficient can furthermore cause the amplification of these perturbations.

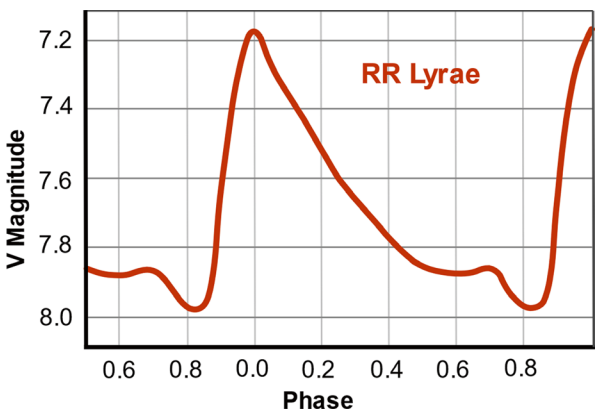


Fig. 5.64 Light curve of RR-Lyra stars (From R. J. Hall commons. https://www.wikimedia.org/)

5.9.2 When Stars Steal Mass

We regard in Fig. 5.65 two masses M_1 and M_2 of a close binary (Sect. 5.2) which circulate with the angular velocity ω around an axis through the center of mass. The radii of the two masses are R_1 and R_2 and their distance from the center of mass are a_1 and a_2 ($a_1/a_2 = M_2/M_1$). The condition for stable orbits: "attractive gravitational force = -repulsive centrifugal force" yields the equation for the angular velocity ω

$$\omega^2 = G \cdot (M_1 + M_2)/a^3 \tag{5.126}$$

where $a = a_1 + a_2$ is the distance between the centers of M_1 and M_2. We regard a small mass m at the distance r_i from the center of mass of M_i. On m acts the attractive gravitational forces by M_1 and M_2 and the repulsive force $F_r = m\omega^2 d_1$, where d_1 is the distance of m from the rotational axis which is the axis through the center of mass of M_1 and M_2. The total force acting on m is then

$$F = -mG\frac{M_1}{r_1^2} + mG\frac{M_2}{r_2^2} + m\omega^2 d_1. \tag{5.127}$$

For F<0 the mass m is attracted towards M_1 for F>0 towards M_2. For F=0 the mass is at an labile equilibrium.

If the mass is located on the straight line between the centers of M_1 and M_2 (Fig. 5.66) the point of labile equilibrium is the Lagrange point L_1 in Fig. 5.66a. Inserting.

$$a_1 = a_2(M_2/M_1) \rightarrow a_1 = a(1 - M_1/M_2);$$

$$d_1 = a_1 - R_1 \tag{5.128}$$

one gets after a longer calculation the maximum distance d_R between M_1 and M_2 where for all distances $d<d_R$ the mass from M_1 is pulled over to M_2. The distance d_R is the **Roche limit**

$$d_R = R_2\left(\frac{M_1}{M_1 + M_2}\right)^{1/3}. \tag{5.129}$$

Definition The Roche limit is the closest distance which a satellite can approach its planet without being pulled apart by the gravitational forces between planet and the different parts of the satellite.

The interface between the two cases where m is attracted by M_1 or by M_2 forms an equipotential surface (black curve in Fig. 5.66) In order to calculate the potential surfaces we choose a coordinate system with its origin at the center of mass M_1 which rotates with the binary system.

For an arbitrary point (x, y, z) with the distance s from the rotation axis through the center of mass S (Fig. 5.66b) and with the distances r_1 and r_2 from the centers of the two masses M_1 and M_2 the equipotential surface is

$$\phi(x, y, z) = -G \cdot \frac{M_1}{r_1} - G \cdot \frac{M_2}{r_2} - \frac{1}{2}\omega^2 \cdot s. \tag{5.129}$$

According to Fig. 5.66a the distance s can be expressed as

$$s^2 = (x - a_1)^2 + y^2 = \left(x - \frac{aM_2}{M_1 + M_2}\right)^2 + y^2. \tag{5.130}$$

In Fig. 5.66b an x–z-cut through the equipotential surface is illustrated. Close to the masses M_1 and M_2 the equipotential surfaces are approximately spheres around the masses as for two separated masses M_i. This is illustrated in Fig. 5.66b by the red circles. Farther away from M_i the equipotential surfaces enclose both masses. Between these two cases the equipotential surface ϕ = 0 is shown in black, where the two equipotential surfaces of the two masses M_i intersect in the Lagrangian point L_1. This surface if called Roche-surface and the volume enclosed by this surface is

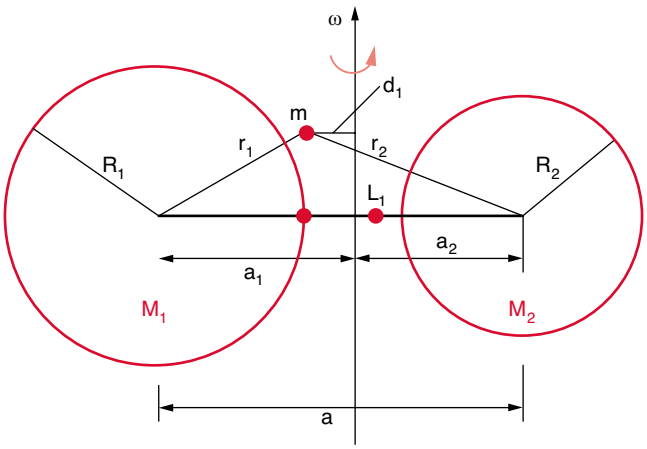


Fig. 5.65 Determination of the Roche limit

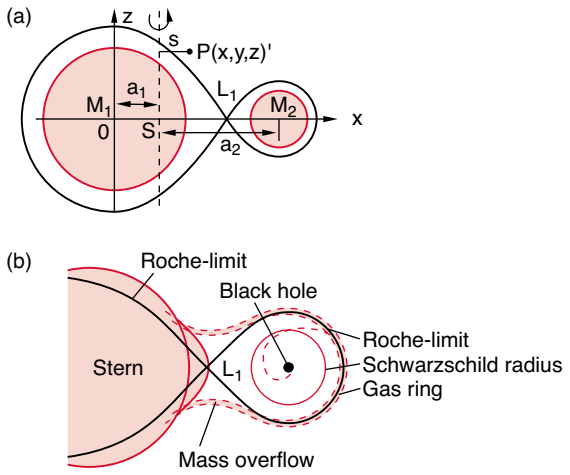


Fig. 5.66 a, b

the Roche-volume. Particles within the Roche volume stay within their mass M_i. because they are more attracted by M_i. Particles outside the Roche-surface can pass from one mass to the other.

If now one of the two components of the binary (e.g. M_1) develops into a red giant star, part of its volume can extend over the Roche surface. That part of its mass, which extends beyond the Roche surface will overflow to M_2. This means that M_1 looses mass while M_2 gains mass. The star with the mass M_2 which steals mass from M_1 can be a star on the main sequence, a white dwarf, a neutron star of even a black hole.

If M_2 is a black hole, the mass stolen from M_1 will be accelerated on a helical path (because the binary system is rotating) into the black hole. The particles, accelerated up to relativistic velocities, will radiate X-rays which can be detected as long as the emitting particles are still outside the Schwarzschild radius.

5.9.3 Novae

Novae (= new stars) have been observed already in medieval times as the sudden flash of a "new" star at a position where before no star could be seen with the naked eye. Such new stars were visible for some nights as bright objects, became slowly fainter and finally disappeared. It is interesting that such events that were observed also in Europe, had not been registered here, because the mandatory opinion of the church attributed the sky as the creation of God, which should be eternal and could not show any changes in time. Therefore the only written documents about novae come from the Arabic and Asiatic regions.

In Fig. 5.67 a typical light curve is schematically shown. The increase of the brightness is very fast, generally within one day and amounts to 7–15 magnitudes ($\approx 10^3$–10^6 times), while the decrease is much slower and is superimposed by irregular oscillations.

The spectral lines show a large Doppler-shift towards the blue, which indicates large radial velocities of the emitting star envelope between 1.000 and 3.000 km/s. These experimental facts indicate, that the nova is a star with fast expanding outer shell. The velocity lies above the escape velocity, which ensures that the emitted particles cannot fall back onto the star. The star ejects its envelope but looses in this process only about 10^{-4} to 10^{-3} of its mass.

The cause of this eruption has been a matter of intense discussion, but can now be explained according to the accepted model as follows:

All observations up to now show that a nova is one component A(M_1) of a binary, which is a star on the main sequence, while the other component B is a white dwarf or a neutron star. When A leaves the main sequence, it enters

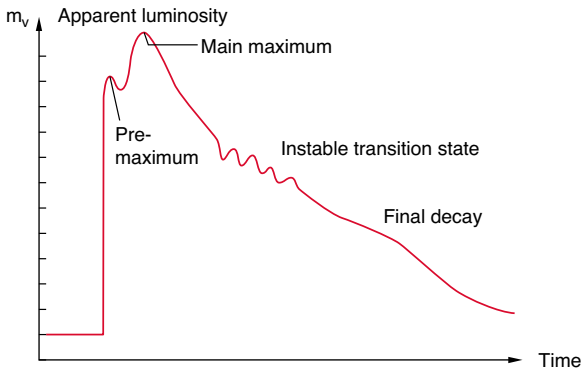


Fig. 5.67 Schematic light curve of a Nova

the giant star stage and expands. As soon as its radius extends over the Roche limit, part of its envelope (which is mainly hydrogen) flows over to M_2. It hits the solid surface of M_2 and heats it up. As soon as sufficient mass has hit the surface of M_2 temperature and density of the hydrogen layer is high enough to ignite hydrogen fusion. Since the matter of the white dwarf is degenerate (see Sect. 5.5) it will be not much compressed by the impact. This results in a particularly large temperature rise because the released energy is not transferred into potential energy but mainly into kinetic energy of the matter at the surface of the white dwarf. Therefore the fusion occurs explosively. The energy released at this explosion expands and ejects the newly acquired envelope. This increases the surface and with it the luminosity.

The sudden increase of the luminosity of a nova is caused by the increase of surface and temperature.

The total energy output of a nova emitted during the time intervall of a few weeks ranges from 10^{37} to 10^{38} J. This corresponds to the energy our sun emits during the time interval of 10^4–10^5 years.

This energy eruption generally does not much affect the companion M_2 of the binary system. This implies that further on mass flows from M_1 to M_2 and the nova-eruption can be repeated.

The ejected gas envelope can form circular nebula (Fig. 5.68).

5.9.4 Supernovae

More seldom than novae (about 20 within 1.000 years in our galaxy), but very impressive are star explosions with a sudden increase of the luminosity by a factor of 10^8–10^{10} are so-called **supernovae.** For about one week a supernova can exceed the luminosity of the whole galaxy where it is located. In Table 5.9 the supernovae, observed in our galaxy and recorded during the last 2.000 years are listed together with their apparent luminosity increase.

Table. 5.9 Historical supernovae

Year	Observed in	Apparent luminosity increase
185	Constellation Zentaur	−6 mag
386	Constellation Schütze	+1.5
393	Constellation Skorpion	−3 mag
1006	Constellation Wolf	−7.5 ± 0.4
1054	Constellation Stier	−6 mag
1181	Constellation Kassiopeia	−2 mag
1572	Constellation Kassiopeia	−4 mag
1604	Constellation Schlangenträger	−2 mag
1680	Constellation Kassiopeia	+6 mag
1885	Andromeda Galaxy	+6 mag
1979	Galaxy Messier 100	+11.6 mag
1987	Large Magelanian Cloud	+3 mag
2014	Galaxy Messier 82	+10.5 mag

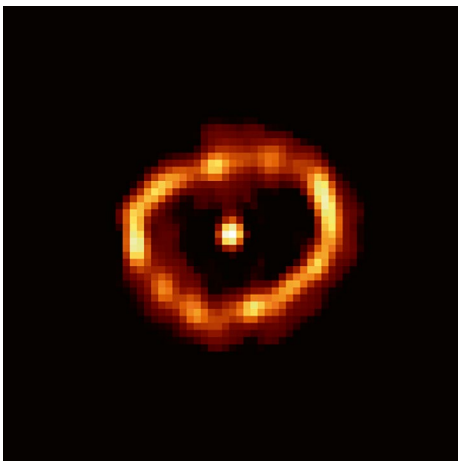


Fig. 5.68 Nebula generated by the ejection of the envelope of Nova cygni (http://hubblesite.org/newscenter)

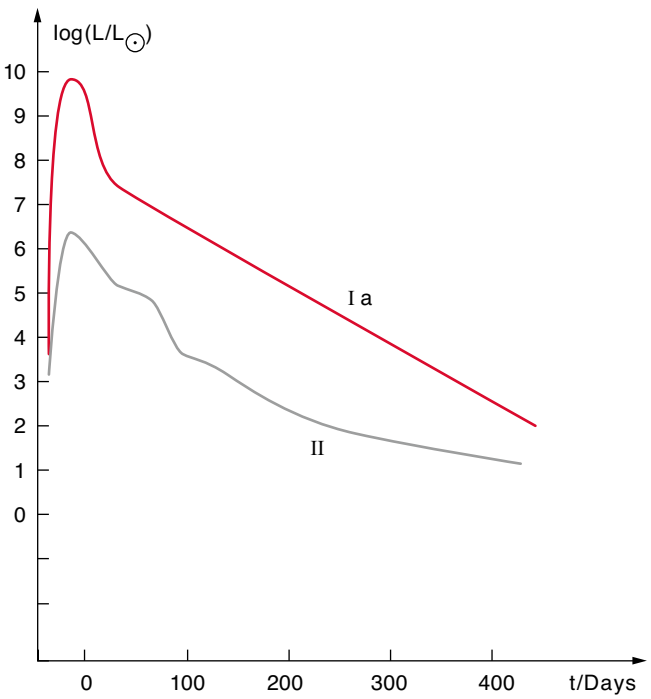
Fig. 5.69 Light curves $\log(L(t)/L_{\odot})$ in units of the sun luminosity for supernovae of type Ia and II

With modern telescopes nowadays more then 100 supernovae are detected per year, because it is now possible to discover supernovae even in far distant galaxies. They are labelled as SN with the year of their detection. If more than one SN is detected in one year they are denoted chronological by capital letters. For instance the SN discovered in 1987 in the Large Magelanic Cloud got the label SN1987A, because it was the first SN detected in this year.

Depending on the observed light curve and the spectra of the SN one distinguishes between two types SN I and SN II (Fig. 5.69).

While for SN II hydrogen lines appear in the spectrum, they are missing for SN I. The two types have different causes, which will be discussed later on. For the SN I a further distinction is made between SN Ia, SN Ib and SN Ic. The large distinction in the spectra of the different SN types indicates different formation conditions.

It turns out that the SN II are caused by the explosion of a massive star ($M > 8\,M_{\odot}$) at the end of ist development. In Sect. 5.5.x we had discussed that this star consists at the end of its life of different onion shell structures where different nuclear fusion processes occur (Fig. 5.41). In the outer shell hydrogen burning takes place while in the inner shells heavy elements like iron Fe or nickel Ni are produced by fusion processes. After these fusion processes extinguish, the central part of the star collapses within less than a second and forms a neutron star or a black hole depending on its mass. The potential energy released during this collapse is partly transferred onto the outer shell (e.g. through shock waves, radiation pressure and neutrinos) and pushes the outer envelope radially outwards with velocities up to 10^4km/s.



The collapse of the core increases its temperature so much that γ-radiation is generated, which can dissolve part of the iron nuclei. This releases protons and neutrons. Because of the large binding energy of the iron nuclei sufficient energy is necessary to start this process. By collisions between protons and free electrons neutrons and neutrinos are produced according to the scheme

$$p^+ + e^- \rightarrow n + \nu_e. \qquad (5.131)$$

The collapsed core has such a high density that only the degenerate pressure of the neutrons can stabilize it against the gravitational pressure (see Sect. 5.7.4) and a neutron star is created.

The star mass from the outer shells falls onto the incompressible core, is reflected at its surface and runs outward as shock wave. There it hits the mass falling down from the outer surface. This causes extremely high temperatures in the shock front. The neutrinos escaping with the velocity of light out of the core are scattered by this shock front with its high density and act as additional acceleration source for the shock front. The outer star envelope is heated up by the shock front from 25.000 K to about 1 Million degrees and the radius of the star increases within one day from some $R_\odot$ up to 10^9 km. It therefore increases by the factor 10^3–10^4. The luminosity of the star increases alone by the temperature increase $\Delta T = 40 \cdot T_0$ by the factor $40^4 = 3 \cdot 10^6$. In addition the increase of the luminosity due to the increase of the surface is 10^6–10^8. The total increase of the luminosity can amount to the factor 10^{10} and it reaches at the maximum of the light curve the value of $10^{10} L_\odot$, which means up to 10 billion times the luminosity of our sun. The resulting luminosity of the supernova can therefore exceed that of the whole galaxy. The total energy emitted during the time of the SN is about 10^{46} where most of the visible radiation is emitted during the first days after the explosion.

The mass of the ejected matter amounts to 0.1–1 $M_\odot$. Its kinetic energy can be calculated from its mass and the ejection velocity as $E_{kin} = 10^{46}$ J. This corresponds to 100 times the energy our sun emits during its whole life.

The by far largest part of the emitted energy (99%) is taken away by the neutrinos. They are produced during the formation of the neutron star in the reaction.

$$p^+ + e^- \rightarrow n + \nu_e.$$

And by dissociation of heavy nuclei by γ-quanta of the radiation at the extremely high temperatures and by collisions induced by the shock front.

Neutrinos play an essential role for the progression during the development of the supernova. Neutrinos produced in the central region of the star are partly absorbed and contribute to the heating of the core. The major part leaves the core and flies outwards where it hits the shock front. Without the neutrinos the shock wave would be soon slowed down by collisions with atoms in the dense gas and would be completely stopped before it reaches the surface of the star. Only the pressure induced by the neutrinos which occurs by the collisions of the fast neutrinos with the atoms in the dense gas, accelerates the shock front and accomplishes that it can reach the edge of the star, where it contributes to the ejection of the star envelope. In spite of the very small cross section for collisions of neutrinos with gas atoms the gas density in the shock front is so large that many collisions occur.

It takes about 1 h before the shock front reaches the surface of the star where it contributes to the increase of the luminosity. Since the neutrinos travel with the velocity of light, they reach the surface of the star one hour earlier. The neutrino signal is therefore detected on earth one hour before the electro-magnetic signal.

The contribution of the neutrinos to the total energy output amounts to 99%. It is therefore larger by a factor 100 than the total electro-magnetic energy integrated over all wavelengths!!.

The temperature in the shock wave rises up to 10^9 K. Passing through the silicon and sulphur shell of the star it causes the fusion of the nuclei into higher elements up to the iron group, particular into the radioactive ^{56}Ni nucleus, which decays with a half lifetime of 6.1d into ^{56}Co. The energy released by the decay heats the outer shells of the star and causes the maximum of the luminosity during the supernova explosion. The decay of ^{56}Co (decay time is 77.1 d) into the stable isotope ^{56}Fe results in an exponentially decaying energy source and is responsible for the exponential decay curve of the supernova.

After several months up to some years the gas density of the expanding star shell has become so low that it is now transparent. It becomes visible as a spherical luminous nebula with the neutron star in its center.

One example it the supernova observed 1054 in the constellation Taurus, which is now visible as Crab nebula (Fig. 5.50). In its center indeed a pulsar as rotating neutron star has been found.

A recent example of a supernova of type II is the supernova 1987A which was discovered in the small Magellan Cloud, 160.000 light years away. Here the neutrinos as well as the visible and X-ray emission has been detected.

Supernova of Type Ia

Supernovae Ia appear like nova-explosions in narrow binary systems, which consist, for instance, of a white dwarf and a close giant star. It is now assumed that like in nova the mass overflow from the giant star to the white dwarf causes a mass-and temperature increase on the surface of the white dwarf (Fig. 5.70). If the mass increase excedes the critical Chandrasekhar limit the star collapses (see Sect. 5.5.2). A

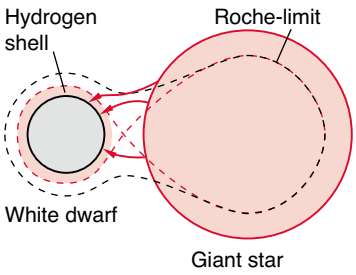


Fig. 5.70 Possible mechanism of the origin of a supernova Ia

mass increase of 0.1 earth masses per year would be sufficient to reach a temperature increase which induces the fusion of hydrogen into helium via the CNO cycle.

The energy released by this fusion process increases further the temperature in the He-layer, inducing the 3α-process

$$\begin{aligned} {}^4_2\mathrm{He} + {}^4_2\mathrm{He} &\rightarrow {}^8_4\mathrm{Be} + \gamma \\ {}^8_4\mathrm{Be} + {}^4_2\mathrm{He} &\rightarrow {}^{12}_6\mathrm{C} + \gamma \end{aligned} \tag{5.132}$$

To produce carbon nuclei. This process proceeds explosion-like and leads to a temperature rise up to 10^9K. At this high temperature the higher fusion processes (i.e. carbon fuses further to heavier elements) start abruptly. This increases the temperature still more. Since the white dwarf consist of degenerate matter (see Sect. 5.5) the temperature increase does not lead to an expansion of the star but the enormous energy density results in expansion forces which surpass the gravitational attraction. The star explodes and is completely teared up.

Such supernovae Ia are the by far brightest SN events. The reason is that here the total star explodes while in SNII only the star envelope is ejected.

The energy released in a SNIa depends on the known Chandrasekhar mass and the temperature. This allows the calculation of the total energy. SNIa can be therefore used as "distance candles" (see Sect. 6.5.10) for the determination of far distances. For SNII the star mass is generally not known. The total released energy can be therefore not calculated.

For SNIa the most important energy source after the explosion is the decay of ^{56}Ni according to the decay scheme

$$^{56}\mathrm{Ni} \overset{6.1\mathrm{d}}{\rightarrow} {}^{56}\mathrm{Co} \overset{77.1\mathrm{d}}{\rightarrow} {}^{56}\mathrm{Fe}. \tag{5.133}$$

The half life time of these processes determines the decay curve of the luminosity.

The similar initial conditions and the equal radio-active decay explain the similar light curves of all SNIa. Since white dwarfs can be found in all galaxies, SNIa can be detected all over the universe.

Besides the SNIa there are other slightly different types which are named SNIb and SNIc. In Fig. 5.71 the classification criteria for the different SN are schematically compiled.

During a supernova explosion exceptional conditions prevail in the star which allow to partly dissociate nuclei that had been synthesized before by nuclear fusion processes. At the dissociation neutrons are released which can be attached to non-dissociated nuclei, thus building up elements heavier than iron (see Sect. 7.1).

Besides their production by slow neutron capture in massive stars supernova explosions are the only occasions where heavier elements can be produced by fast neutron capture. The fact, that we find on earth such elements proves that at least part of the matter we are built of, comes from exploding stars. Such supernovae explosions not only have produced the heavy elements but have also enabled the formation of our solar system by the shock waves inducing the instability of the gas cloud that has formed our solar system.

5.9.5 Planetary Nebula and Supernova Remnants

In the foregoing sections it has been shown that a star passes, after leaving the main sequence, several instable phases, which can lead to the ejection of a part of its envelope. For instance during the red giant state a slow continuous flow of gas into space occurs (star wind). Because of the large star diameter the gravitational attraction is smaller and cannot prevent hot atoms at the surface from escaping into space. The velocity of the out streaming gas is about 20–40 km/s. The resulting mass loss amounts to about $10^{-8}M$ per year.

Besides this slow continuous gas loss also a short time ejection of matter can occur during short unstable phases. This happens, for instance, when at the end of the giant star phase the star develops to a variable star with thermal oscillations. These cause an ejection of fast particles which overtakes the slow star wind and the superposition generates a shock front which forms a planetary nebula. The central star, which is generally a white dwarf, excites by its UV radiation the atoms in the nebula and induces the nebula to shine. One can see a luminescent ring nebula surrounding the central star, which expands with a velocity of about 20 km/s. The total mass of the nebula amounts to about 0.1–0.3 M of the mass of the central star. One example of such a planetary nebula is the ring nebula in the constellation Lyra (Fig. 5.36) and the ring nebula of the supernova 1987A (Fig. 5.72).

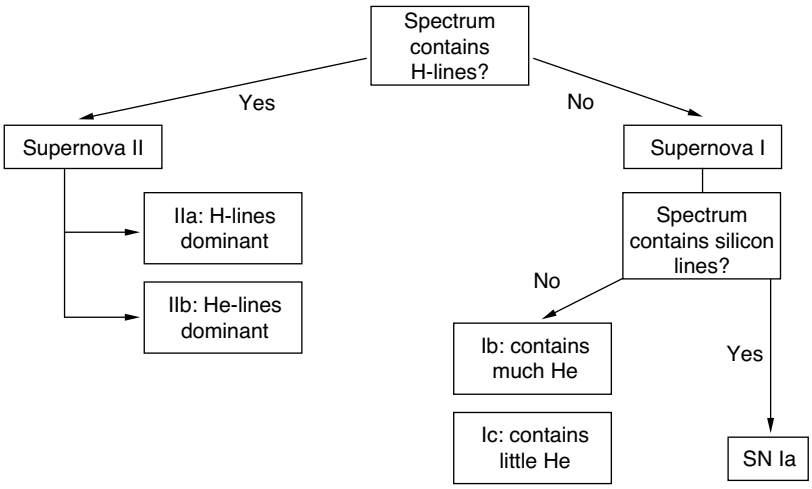


Fig. 5.71 Classification criteria for the different types of supernovae (Max Camencind Heidelberg)

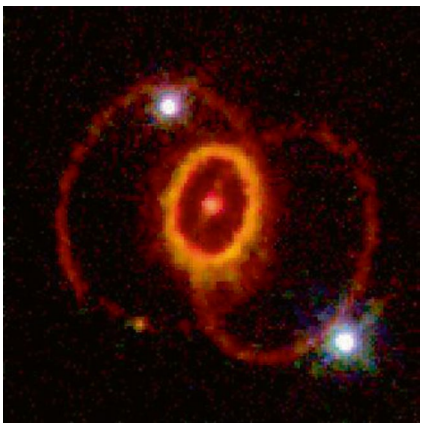


Fig. 5.72 Ring-nebula as remnant of the supernova 1987A (Photo of the Hubble space telescope NASA)

Planetary nebula are formed at the end of stars with masses $0.8\,M_\odot < M < 8\,M_\odot$. Typical lifetimes between their formation and their end by recombination of the ions and electrons in the plasma are about 10.000 years.

Up to now about 3000 planetary nebulae have been discovered. It is, however, estimated that alone in our galaxy more than 50.000 of such nebulas exist.

Besides these nebulae huge gas clouds are formed by novae and supernovae explosions. In nova-explosions the particles are ejected with velocities up to 3.000 km/s. Since these velocities are larger than the escape velocity the particles do not fall back onto the central star but escape into the interstellar space. During the expansion the gas cools down and small particles can be formed by condensation. They scatter and absorb the light from the central star. These gas and dust-masses form the dark part of the interstellar clouds (Fig. 5.x).

A particular spectacular gas outburst occurs in supernova explosions, where the envelope of the star is ejected with velocities up to 10.000 km/s. The star diameter increases during one day from sun size to the diameter of the whole solar system.

After some years the density of the ejected matter has decreased so much, that it becomes transparent and appears as luminous nebula. It represents the visible gaseous remnant of the supernova. By collisions with the atoms and molecules of the surrounding interstellar gas the temperature rises up to 10^6K. This hot gas emits besides visible and UV radiation also X-rays.

The supernova remnants appear as spherical shell which expands with high velocity. The radial velocity can be determined from the Doppler-shift of the spectral lines. Observed from the earth this shell appears as concentric ring around the central star. which is generally a rotating neutron star that appears as pulsar.

Measuring at the time t_1 after the supernova explosion the angular ring diameter $2\alpha = 2R/r$ the absolute distance r can be obtained from the calculated diameter $R(t) = \int v(t)dt$. Therefore supernovae represent good distance indicators for large distances where other methods fail.

5.10 Summarizing the Star Development

We will in this section shortly recapitulate, illustrated by Figs. 5.73, 5.74 and 5.75 the most important development phases of stars and its dependence on the mass of the star [8].

Stars are born as associations by contraction of huge cold gas- and dust-clouds of low temperature, low density but large total mass. During the contraction they fragmentate and form rotating partial clouds, which have an oblate shape caused by centrifugal forces, axial symmetric around the angular momentum direction. In their center protostars

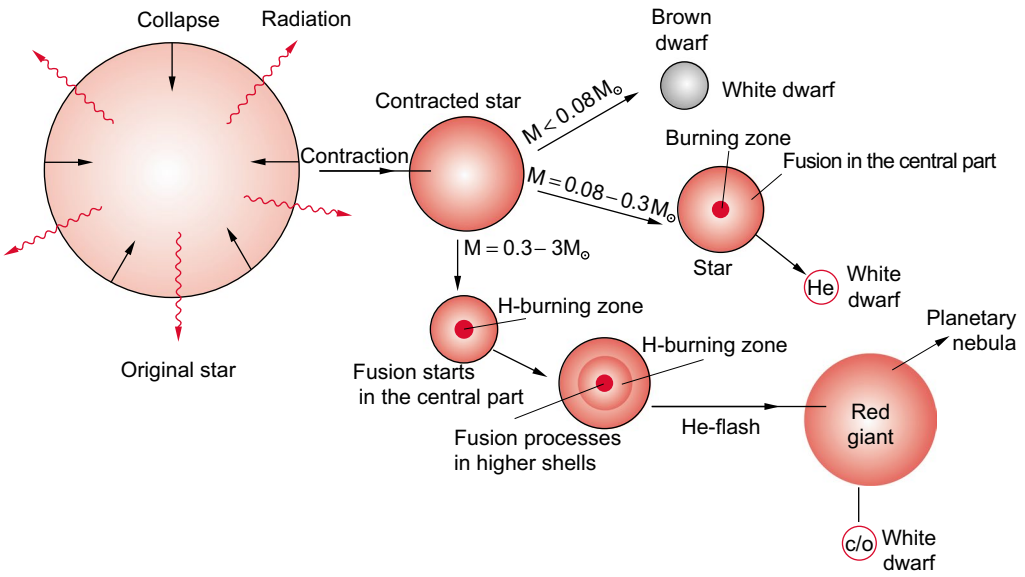


Fig. 5.73 Schematic representation of the development of stars with small to medium masses

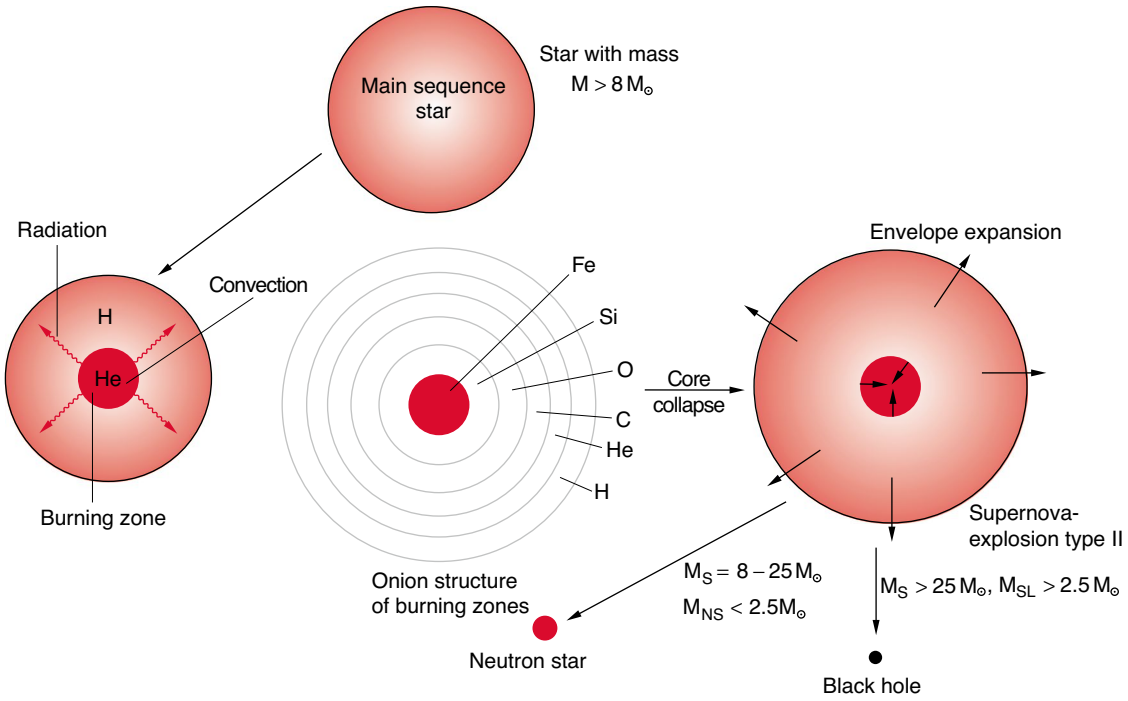


Fig. 5.74 Development of massive stars

are formed where the matter from the outer regions falls onto the protostar.

The further development of the protostar depends critically on the mass. For $M < 0.08\,M_\odot$ the temperature in the center of the protostar does not reach the temperature for hydrogen fusion. The star contract and slowly cools down and becomes a cold planet-like **brown dwarf**.

For stars with a mass $M > 0.08\,M_\odot$ the temperature in the center increases above the critical value for hydrogen fusion ($T > 4 \cdot 10^6$K) Due to its higher temperature the thermal pressure compensates the gravitational attraction and the star starts a stable phase on the main sequence in the HRD.

Stars with a mass $0.08\,M_\odot < M < 0.3\,M_\odot$ have only a small central burning zone. They are fully convective, i.e.

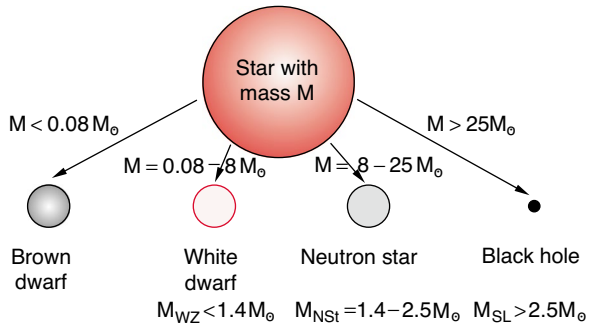


Fig. 5.75 Schematic diagram for the final states of stars with different masses M

the hydrogen outside the burning zone can be transferred by convection into the burning zone and can there be fusioned to helium. Therefore these stars can consume their whole hydrogen supply. They have a long lifetime and become finally white dwarfs (Fig. 5.73).

In stars with $0.3 < M < 2.3\,M_\odot$ a larger central region can reach fusion temperatures. The temperature gradient in this larger range is therefore small and the energy transport occurs mainly by radiative transfer. In the outer layers convection takes over. At the end of the main sequence development the central region consists mainly of helium while in the surrounding shells hydrogen burning goes on. Because the hydrogen fusion increases the temperature, the envelope of the star expands and the star becomes a red giant. Since the helium in the central core cannot yet fuse the core contracts which increases the temperature until the fusion of He to carbon ignites (3α-process). Since the matter in the core is degenerate the temperature increase does not lead to an expansion of the core, but the released energy results in an explosion-like temperature increase which is transferred to the outer shells and leads to a sudden increase of the luminosity (helium flash). The temperature increase of the star envelope leads furthermore to the ejection of an essential part of the star mass, generating a planetary shining nebula, which is excited by the radiation of the central rest star (essentially the hot core).

Stars in the mass range $2.3\,M_\odot < M < 8\,M_\odot$ reach after the He-burning sufficiently high temperature to ignite the CNO-fusion cycle. The temperature increased by this process starts the fusion of all heavier elements up to iron. The core of the star is convective (because its temperature gradient is large) while the outer layers show a radiative energy transfer. Due to a large mass loss by star winds and the massive ejection of the outer layers which forms planetary nebula, the mass of the rest star is not large enough to form a neutron star. The final state is therefore, like for the foregoing mass group, a white dwarf and a planetary nebula (Fig. 5.74). The white dwarf now contains more heavy elements up to iron.

In the core of massive stars with masses $M > 8\,M_\odot$ nearly all lighter elements experience fusion up to iron. In the shells around this iron core the temperature decrease with increasing distance r from the center. Here different fusion processes take place which produce all lighter elements (see Fig. 5.74).The shell structure of the star is like an onion shell, During these fusion processes a large part of the outer shells is ejected while the essential part of the star mass is concentrated in the core. As soon as the limiting Opopoenheimer-Voolkov mass is exceeded, the core collapses and becomes for $M < 25\,M_\odot$ a neutron star, for $M > 25\,M_\odot$ a black hole. The huge energy released during this collaps results in a supernova of type II explosion.

In Fig. 5.75 the different final states of stars with different masses of the initial star are compiled.

Note: The masses of the final states are much smaller than the initial masses. This has two reasons:

(a) During the unstable end phases large masses are ejected into the surrounding space which results in an essential mass loss of the star.
(b) The gravitational energy of the final states has a much larger negative value because of the much smaller radius. This results according to (5.89) to a relative mass loss

$$\Delta M/M_\odot = -(3/5)GM_\odot/\left(Rc^2\right)$$

Note: In the classification of the different mass ranges the initial star masses are used.

5.11 Thought Provoking

We have learned in this chapter that stars start their life as cold molecular hydrogen clouds, which contract due to the gravitational attraction. At the beginning of their path on the main sequence they consist mainly of protons and electrons. During their main sequence development and after having left the main sequence they synthesize by nuclear fusion heavy elements. The released fusion energy (nuclear binding energy) is transferred to the emitted radiation. The mass defect by the nuclear fusion amounts to about 1% of the star mass. For our sun this corresponds to a radiation energy of

$$W = 3.9 \cdot 10^{26}\mathrm{W} \cdot 10^{10}\mathrm{a} \approx 10^{44}\mathrm{J}.$$

This radiative energy is produced by fused nuclei, which represent about 10% of the total star mass.

At the contraction of the star core to the radius $R \approx 10^4$km of a white dwarf, which takes only a few minutes, the released gravitational energy amounts to 10^{43} J which is about 10% of the total radiation energy of the star emitted over the lifetime of the star of 10^{10} years. This shows already that the gravitational energy gives an essential contribution to the total emitted energy.

This becomes even more impressive for massive stars, where the final state is a neutron star. Here the mass of the central region of the star is converted into neutrons, while the envelope of the star still mainly consists of hydrogen.

In an exaggerated formulation the star development from its birth to its death consists in a transformation of protons and electrons into neutrons.

Now the salient point comes: The mass of a neutron is larger than the mass of proton + electron, which implies that energy is necessary to induce the reaction

$$p^+ + e^- \rightarrow n + \nu_e$$

For the formation of neutron-rich nuclei in the shell of neutron stars again energy is needed.

Where does the huge amount of energy comes from, which the star emits during its whole lifetime (several billion years)???

Of course part of this energy comes from nuclear fusion as has been outlined in this chapter. However, the fusion becomes only possible due to the temperature increase caused by the gravitational forces that contracted the gas cloud.

The reader, educated in Physics, certainly believes in the energy conservation law. Therefore a physically plausible source of energy must be present, that can supply these enormous energy. The answer is: It is the gravitation, which we always have regarded as the weakest of all 4 forces, known in Physics. The gain of gravitational energy at the formation of a neutron star is much larger than the energy won by nuclear fusion during the whole lifetime of the star. This is even more pronounced if the final state of the star is a black hole.

For all stars which end as neutron stars or black holes we can make the following energy balance:

In spite of the various and complex nuclear fusion processes, which we often regard as the most effective energy production, the true energy source for the whole life of a star is the gravitational energy. The nuclear energy processes have delayed the collapse of the star. They help to elongate the time interval for the energy production by fusion processes and prevent, that the energy release occurs in a short time leading to an explosion. For the energy release during the whole lifetime of a star their contribution is marginal.

This is a surprising result!

For this energy consideration we still have neglected the matter, ejected into space during Nova-or Supernova explosions. The kinetic energy of this ejected matter also contributes to the total energy release of a star. After the cooling this ejected matter forms molecular clouds which are the birth places of new stars. The by far major part of the energy released during a supernova explosion is, however, carried away by the neutrinos.

Summary

- A star is characterized by its mass, its radius, its luminosity, its surface temperature and its spectral type.
- The Hertzsprung-Russel diagram (HRD) plots the luminosity of stars against their surface temperature. Stars stay during their stable burning phase on a limited range of the HRD, called the main sequence, because most of the observable stars are located in this range.
- There is a relation $L \propto M^\alpha$ ($\alpha = 3$–4) between the mass M of a star and its luminosity L.
- Stars are born by contraction of huge cool molecular and dust clouds, which become gravitational instable, if the mean kinetic energy of the cloud particles is smaller than the amount of the negative gravitational energy (Jeans-Criterion).
- Stars are generally born in star associations not as single stars.
- The collaps of a rotating cloud does not occur uniformly, but is governed by the transport of angular momentum from the inner to the outer part of the collapsing cloud. Furthermore the velocity of the collaps is influenced by the dissociation of the molecules, the ionization of the atoms and by the increasing radiation pressure.
- At first protostars are formed which radiate in the infrared, but have not yet reached sufficiently high temperatures to ignite nuclear fusion processes.
- As soon as the ignition temperature for fusion processes has been reached (this is realized for stars with masses $M > 0.08\,M_\odot$) the pressure increase caused by the higher temperature, stops the collaps completely. The stable

phase of the star development starts (main sequence state).

- The time on the main sequence and the phases after having left it, depends critically on the mass of the star. The larger its mass the shorter is the lifetime of the star.
- Stars with $M < 0.26\,M_\odot$ contract, just after the H-fusion has ended, to white dwarfs. Stars with $M > 0.26\,M_\odot$ pass after having left the main sequence through several instable phases where the central part contracts while the star envelope increases and the radius oscillates.
- The final state of stars is, depending on the mass, a white dwarf ($M < 1.4\,M_\odot$), a neutron star ($1.4\,M_\odot < M < 3\,M_\odot$), or a black hole ($M > 3M_\odot$).
- In white dwarfs the gravitation pressure is compensated by the degenerate pressure (Fermi-pressure) of the electrons.
- In neutron stars the protons + electrons are converted to neutrons. The star is stabilized by the degenerate pressure of the neutrons. Neutron stars posses a strong magnetic field.
- During the formation of a neutron star in the core the star envelope is ejected caused by a shockfront and the neutrino impact.
- The radius of white dwarfs is about 10^4 km, that of neutron stars about 10 km.
- Pulsars are fast rotating neutron stars, where the magnetic dipole axis does not coincide with the rotation axis. They emit synchrotron radiation extending from the X-ray region up to the infrared range.
- Supernovae of type I originate, if in a narrow binary of a white dwarf and a companion, which can be a red giant, matter overflows from the giant to the white dwarf. This brings the white dwarf above the stability limit of the Chandrasekhar mass. It collapses and the resulting sudden increase of its temperature ignites the carbon burning, which finally tears the total star apart.
- Since their energy release can be calculated, supernovae of type Ia can be used as distance candles for measurements of large distances.
- Supernovae of type II originate in massive stars when during the contraction of the core the whole envelope is ejected.

Problems

5.1 The two components of a photometric binary should have equal radii $R_1 = R_2 = R$ but different surface temperatures T_1 and T_2. What is the light curve of the system with a period τ?

5.2 The rotation period of a pulsar is $\tau = 1.2$ms . It increases per year by $\Delta\tau = 10^{-5}\tau$. What is the rotational energy for a mass $M = 2\,M_\odot$ and a radius $R = 8$km when we assume a uniform mass distribution $\varrho(r) = \varrho_0$? How large is the emitted radiation power?

5.3 Calculate the gravitational pressure at the radius R_0 of a homogeneous sphere where $\varrho = \varrho_0 = \text{const.}$ for $r \le R_0$ and $\varrho = \varrho_0 \cdot e^{-ar}$ for $r \ge R_0$.

5.4 What is the red shift in the gravitational field.
1. of the earth at its surface.
2. of a neutron star with $M = 2M_\odot$ and $R = 10^4$m?

5.5 What is the minimum mass of a body with the density ϱ when the gravitational bindings energy becomes larger than the el, magn, binding energy ($\varepsilon = 1$eV per atom pair) for a mean mass number A of the atoms ($\varrho = 4\text{g/cm}^3, A = 50$).

5.6 What is the upper limit for the radius R of planets, if the gravitational pressure should be compensated by repulsive el. magn, forces (compression pressure $p = 3 \cdot 10^{11}$Pa).

5.7 How fast can a neutron star rotate, before a mass at the equator is ejected i.e. the centrifugal force becomes larger than the gravitational force ($M = 2M_\odot, R = 10^4$m)?

5.8 At the 23.02.1987 10 neutrinos were detected within 10 s, coming from the supernova 1987A with a distance of 50 kpc. The detector consisted of a tank filled with 2140 t water. The detection used the Cerenkov radiation of the charged leptons and light nuclei produced by recoil in collisions with the neutrinos. The average energy of the neutrinos was 12 meV. At this energy the collision cross section is $\sigma = 4 \cdot 10^{-46}\text{m}^2$. Determine the neutrino flux ($\text{m}^{-2}\text{s}^{-1}$) and estimate how much energy was released by the supernova within 10 s?

References

1. Gaia - Wikipedia
2. Constellation Guide | StarDate Online
3. Hanbury Brown and Twiss effect - Wikipedia
4. Binary star - Wikipedia Binary Star Systems: Classification and Evolution | Space
5. Harvard Spectral Classification | COSMOS (swin.edu.au)
6. Stellar classification | astronomy | Britannica
7. Protostar - Wikipedia
8. Gaia (spacecraft) - Wikipedia
9. Stellar classification - Wikipedia
10. Phillips, A.C. (1999). The Physics of Stars. John Wiley & Sons
11. Richard B. Larson (10 September 2003). "The physics of star formation"
12. E. A. Bergin, M. Tafalla: Cold Dark Clouds: The Initial Conditions for Star Formation. In: Annual Review of Astronomy & Astrophysics. Bd. 45 (2007), S. 339–396 (bibcode:2007ARA&A..45..339B)

13. St. W. Stahler, F-Palla: The Formation of Stars (Wiley -VCH Weinheim 2004)
14. RF. V. Pound, G. A. Rebka: Appaent weight of Photons. Phys. Rev. Lett. 4, 337 (1960)
15. Abbott, Benjamin P.; et al. (LIGO Scientific Collaboration and Virgo Collaboration) (2016) "Properties of the binary black hole merger GW150914". Physical Review Letters. 116 (24): 241102
16. First observation of gravitational waves - Wikipedia Nova explosion - Bing
17. Black hole - Wikipedia
18. https://en.wikipedia.org/wiki/Quasar
19. Roche limit - Wikipedia

The Development and Present Structure of the Universe

6

Contents

6.1 Experimental Proofs of a Finite Expanding Universe 178
6.2 The Metric of a Curved Space 182
6.3 The Standard Model 182
6.4 Formation and Structure of Galaxies 199
Summary 211
Problems 212
References 212

In former times, it was assumed that the universe was static and spatially infinite. Astronomers favored generally the aspect of a homogeneous and isotropic universe because this satisfied their aesthetic picture of God's creation. The assumption that the universe looks equal for all observers, independent of their location and viewing direction is called the **cosmological principle**.

On a small scale, the universe is, of course, not uniform, because there are planets, stars, gas- and dust clouds, galaxies and galaxy clusters, which shows that the matter is not at all homogeneously distributed [1, 2]. However, on a very large scale (of the order of 10^9ly), the galaxies confined in a volume of $\left(10^9\text{ly}\right)^3$ are uniformly, i.e. statistically distributed (Fig. 6.1).

Einstein introduced into his field equations for the description of the static universe a "**cosmological constant**", in order to prevent the collapse of the universe due to the attractive gravitational force between the galaxies. After he was informed by Edwin Hubble about Hubble's discovery of the expansion of the universe (Fig. 6.2), he declared this introduction of the cosmological constant as his biggest blunder.

Presently, it is, however, discussed again that Einstein's cosmological constant might not be a blunder but might be realistic in order to explain the development of the universe (see Sect. 6.10).

Nowadays, we believe that the universe is neither static nor infinitely extended in space. Also, its homogeneity might not be realistic as more detailed measurements seem to indicate.

The Standard Model of the Universe, which is generally accepted by most astronomers, is based on the following assumptions:

- The universe started about 13–14 Billion years ago from an extremely hot fireball and has expanded ever afterward (Big Bang Theory). According to more detailed measurements, the age of the universe is now determined to be $13.75 \cdot 10^9$ years [3].
  Until recently, it had been assumed that the expansion velocity decreases in the course of time, because of the attractive gravitational force between the galaxies. Recent very accurate distance measurements based on supernovae SNIa, which served as standard candles, showed, however, that the expansion rate accelerates. Causes for this acceleration are presently intensely discussed and are partly attributed to the dark energy (see Sect. 7.2).
- During the first 4 min of the big bang, the three lightest nuclei, namely H, D and He were formed and to a small amount also Li. All the other elements have been synthesized later on by fusion processes in the central part of stars (see Sect. 5.5). The elements heavier than iron were generated during supernovae explosions in the hot expanding star envelope.
- About $3.8 \cdot 10^5$ years after the big bang at the temperature of $T \lesssim 3000$ K, the decoupling of matter and radiation took place. Free electrons could recombine with ions to form neutral atoms and molecules, which have much smaller absorption and scattering coefficients of the spectral continuum of the radiation than free electrons. **The expanding universe becomes transparent**.
- After sufficient cooling of the expanding universe at about 10^7–10^8 years after the big bang, galaxies and stars were born by contraction of huge cold gas- and dust clouds, This formation of new structures at this relatively young age of the universe can be only explained by assuming **dark matter**, with a total mass much larger than that of the visible matter.

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_6

Fig. 6.1 Map of galaxies in a quadratic sky section with 6° sidelength, photographed by Rudniki et al. Krakau Poland. The average distance of the 10.000 galaxies in this picture from the earth is about 2–3 Billion lightyears [1]

Fig. 6.2 Einstein at the Palomar telescope with Edwin Hubble

The question is now: Which experimental facts support these assumptions?

What are the open questions that could not be convincingly answered up to now? [2, 5–9].

6.1 Experimental Proofs of a Finite Expanding Universe

The standard model of Astronomy is supported by several experimental observations which prove the essential points of the Big Bang Theory.

6.1.1 The Expansion of the Universe

The American astronomer Edwin Powell Hubble (1889–1953) discovered in 1929 that the wavelengths of spectral lines emitted from distant galaxies are shifted to the red. Since he had measured already the distances of many galaxies, he noted that the wavelength shift is proportional to the distance of the galaxy (see Sect. 6.3). He interpreted his observation as Doppler shift and concluded that all galaxies recede for the observer with a velocity

$$v = H \cdot r. \tag{6.1}$$

The factor H is the **Hubble parameter**. It is time-dependent. Its present value at time t_0 is the **Hubble constant** $H(t_0) = H_0$.

Measurements with the Hubble space telescope, the space satellite WMAP (Wilkinson Microwave Anisotropy Probe), and in particular the accurate measurements of many far distant galaxies with the supernovae SNIa as standard candles (Sect. 5.8) has proved that the red shifts of even very far distant galaxies ($r > 5 \cdot 10^9$ly) could be measured very precisely. This has improved the accuracy of the Hubble constant considerably. Its value accepted today lies between the limits [2, 10]

$$H_0 = (68 - 75)\text{km/s pro Mega parsec}. \tag{6.2}$$

If the expansion velocity v of the universe decelerates during the expansion, the value $H(t < t_0)$ would have been larger, if it accelerates, $H(t < t_0)$ would have been smaller than the present value H_0.

Note: the values of H_0 obtained by different authors with different methods differ more than the error bars given by the authors. For instance, the value obtained from measurements of far distant galaxies with the X-ray satellite CHANDRA is $H_0 = (77 \pm 4)$km/s pro Megap arsec, while the value obtained from distance measurement of supernovae in distant galaxies is $H_0 = (73.8 \pm 2.4)$ km/s pro Mega parsec. The most accurate value $H_0 = (67.8 \pm 1.0)$ km/s pro Mega parsec has been obtained form measurements of the anisotropy of the cosmic background radiation by the space satellite Planck.

6.1.2 Cosmic Background Radiation

The second essential support of the Big Bang theory was provided by the discovery of Cosmic Background Radiation in 1964 by *Penzias* and *Wilson* (Fig. 6.3). This discovery was fortuitous when the two scientists at Bell Telephone wanted to clear up the unexpected background noise of a microwave antenna that acted as a receiver for the communication with satellites (Fig. 6.4).

Fig. 6.3 A. A Penzias, R. W. Wilson

It turned out that from the universe an anisotropic electro-magnetic background radiation is received with a spectral distribution that corresponds to a thermal Planck radiation at a temperature of $(2.725 \pm 0.002)K$.

It is interpreted as the remnant of the radiation at the time of the big bang, which had been shifted toward longer wavelengths by cooling due to the expansion of the universe (see Sect. 6.3). The existence of this 2.7 K radiation had been predicted *by G. Gamow* who estimated its temperature as $T = 10$ K.

Note: Long-time measurements of the spectral and the angular distribution of the cosmic background radiation by the satellite COBE (cosmic background explorer) have revealed a small anisotropy of the radiation, which varies between 2.7251 K and 2.72507 K depending on the direction (Fig. 6.6).

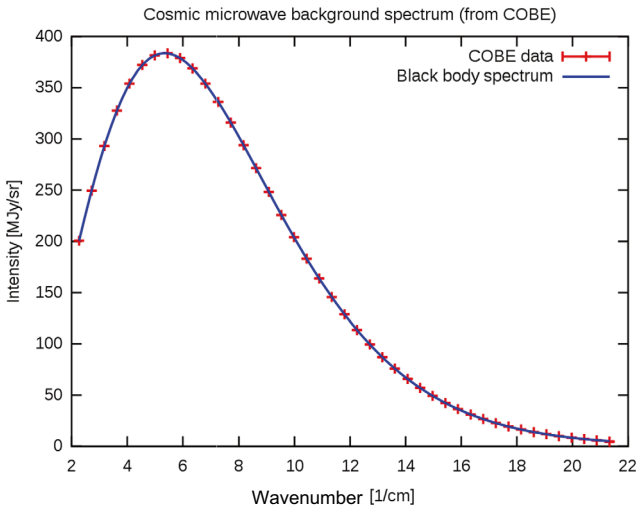


Fig. 6.5 Spectral distribution of the cosmic background radiation. The measurements (red crosses) fit exactly the theoretical curve (blue curve) of the Planck radiation (https://www.bing.com/images/search?q=spectrum+of+cosmic+background)

Much more accurate measurements of this anisotropy were achieved with the satellite WMAP (**W**ilkinson **M**icrowave **A**nisotropy **P**robe) [11]. This new satellite, which is based on previous achievements of *D. T. Wilkinson*, is the prime source of Cosmologists, because it gives detalled information about the universe at the time 380.000 years after the big bang. At this time, the radiation

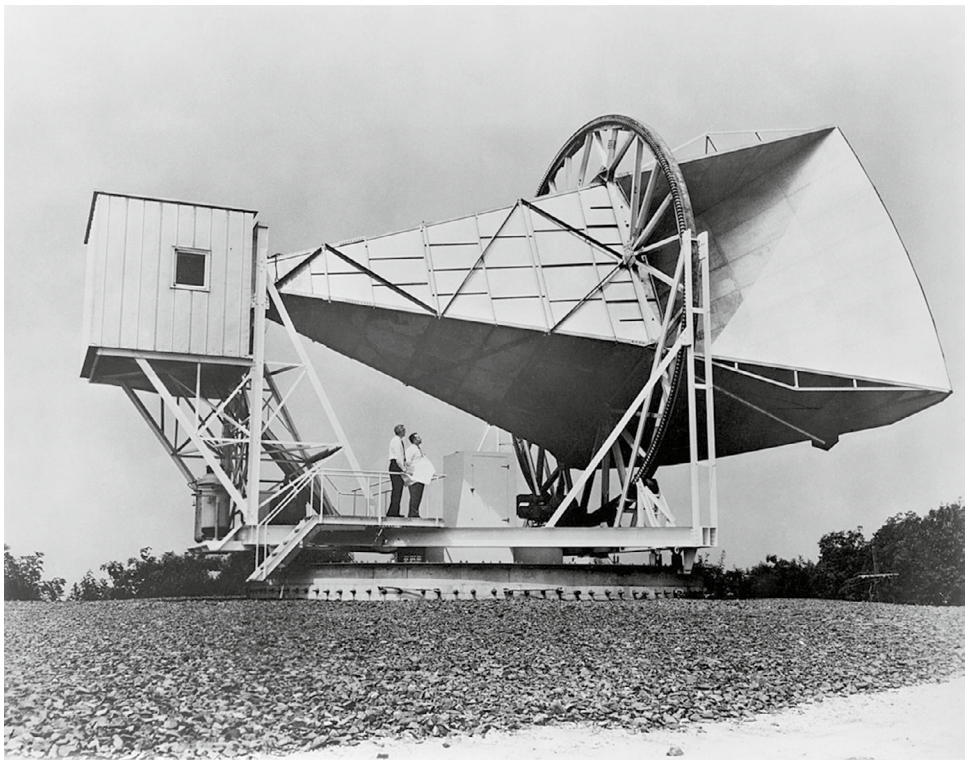


Fig. 6.4 Horn Antenna, with which Penzias and Wilson discovered the cosmic microwave background radiation

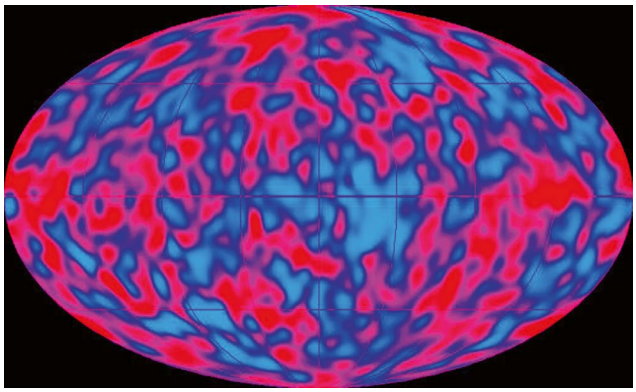


Fig. 6.6 Spatial temperature fluctuations of the cosmic background radiation measured with the satellite KOBE (The KOBE data sets were developed by NASA Goddard Space Flight Center under the guidance of the COBE science working group)

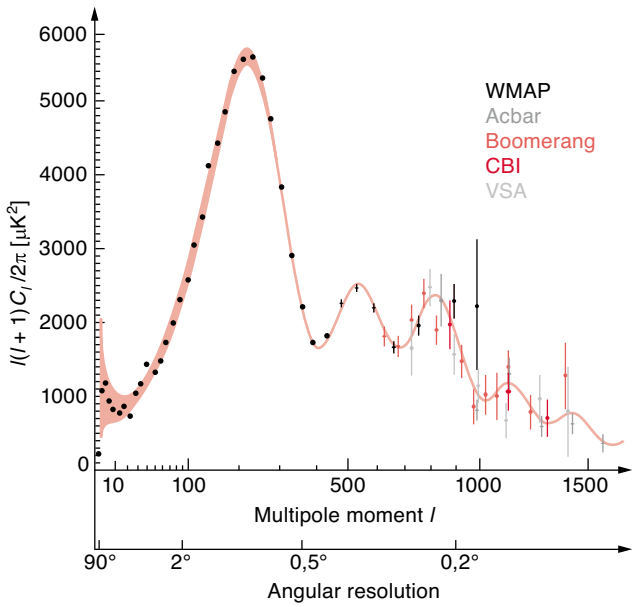


Fig. 6.7 Multipole expansion of the intensity of the cosmic background radiation, measured with the satellite WMAP (wikipedia.org/wiki/File.WMAP2008)

decoupled from matter and the anisotropy of the background radiation reflects the fluctuations of the matter density at this time (see Sect. 6.3.6).

In order to describe this anisotropy in a quantitative way, the weak angular dependent intensity of the background radiation is expanded into a series of spherical harmonics $Y(\ell, m)$ (multipole expansion). In Fig. 6.7, the multipole expansion of the cosmic background radiation is shown, based on all measurements up to now. The higher multipole terms reflect the fluctuations of the background radiation in small angular ranges, i.e. small volumes in space, while the terms with small values of l give the large- scale fluctuations.

In addition, a large-scale anisotropy is found, which reflects the motion of the earth through the universe and contributes to the Doppler effect.

Even more accurate measurements with a better angular resolution could be performed with the satellite Planck of the European Space Agency ESA, which recorded within the time 2009–2012 a wealth of data. Their analysis provided very accurate cosmic parameters such as the age of the universe, the Hubble constant and the matter density in the universe.

6.1.3 Abundance Ratios of the Light Elements

The standard model postulates that during the early time of the universe, only the four lightest elements H, D, He and Li have been generated, while all other elements originate by fusion processes in the central region of stars. The ratios of these elements, measured today by space probes are

$$\mathrm{H} : \mathrm{D} : \mathrm{He} : \mathrm{Li} = 1 : 10^{-5} : 0.082 : 10^{-9}$$

which confirms the ratios predicted by the **Standard Model**. This is a further experimental confirmation of the Standard Model.

6.1.4 The Olbert's Paradoxon

An early hint that the universe filled with stars, must be finite, was already given by *W Olbert* (1758–1840). He answered the question: Why is the night sky after all dark? The simple answer: "because the sun has set down below the horizon" is in fact not obvious, as will be shown below.

In an infinitely extended homogenous universe, the night should not be dark. This can be seen as follows.

We assume the mean density n of stars in a homogeneous infinite universe where every star has the mean luminosity L. The intensity (energy flux density $\mathrm{W/m^2}$) of a star at the distance r is on earth

$$I_1 = L/\left(4\pi r^2\right). \tag{6.3}$$

In a spherical shell between the radii r and $r + \mathrm{d}r$, the number of stars is $N = n \cdot 4\pi r^2 \mathrm{d}r$ The intensity produced by these N stars on earth is

$$\mathrm{d}I = N \cdot I_1 = N \cdot L_1/4\pi r^2 = n \cdot L \cdot \mathrm{d}r. \tag{6.4}$$

All stars with the distance $\mathrm{r} < r_{\max}$ generate the intensity on earth

$$I(\mathrm{r} < r_{\max}) = n \cdot L \int \mathrm{d}r = n \cdot L \cdot r_{\max}. \tag{6.5}$$

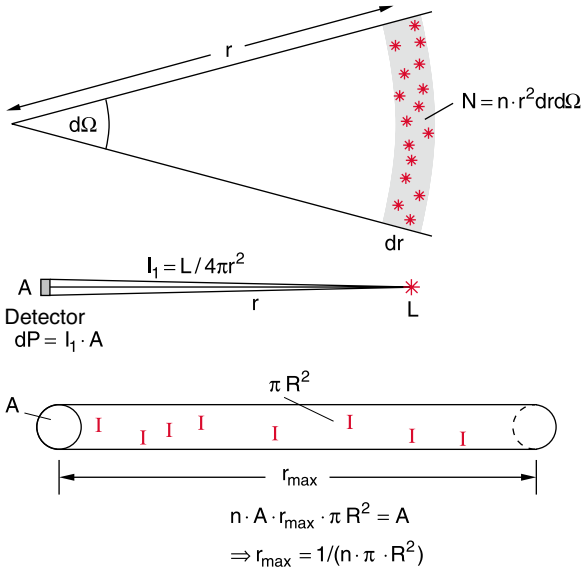


Fig. 6.8 Determination of the maximum distance of the stars for the explanation of Olbert's paradoxon

We have to take into account that the light from a far distance star at $r = r_{max}$ can be concealed by closer stars. With the star radius R, complete covering occurs if the effective area of all stars in a cylinder with cross section A and length r_{max} reaches the value.

$$F_{\text{eff}} = n \cdot A \cdot r_{\max} \cdot \pi r^2 = A \quad r_{\max} = 1/\left(n\pi R^2\right). \quad (6.6)$$

In this case, the total area of the cylinder is covered by stars (Fig. 6.8).

The total intensity of all stars up to $r = r_{max}$ is therefore

$$I(r_{\max}) = n \cdot L \cdot r_{\max} = L/\pi R^2. \quad (6.7)$$

Now the quantity $L/\pi R^2$ gives the surface brightness of a star with radius R (emitted radiation power per unit surface). This means: In every direction towards the sky, the surface brightness should be the same as that of a star. Independent of the direction, If we look into the sky, independent of the direction, we would see a star surface. Since the surface brightness is independent of the distance r, every point of the sky should have the same surface brightness as our sun. This result, which is obviously senseless, is called **Olbert's Paradoxon**.

The Standard Model solves this paradoxon. Because it assumes a finite expanding universe. With the expansion velocity v and the time τ, back to the Big Bang, we obtain the maximum extension r_{max} of the universe

$$\tau = 13.75 \cdot 10^9 ly \rightarrow$$

$$r_{\max} = \int v(t)\mathrm{d}t = 13.75\text{ly}. \quad (6.8)$$

On the other side, the value of r_{max} in Olberts consideration is

$$r_{\max}^{\text{Olb}} = \left(n \cdot \pi R^2\right)^{-1}.$$

Inserting typical numerical values $R = R_\odot = 7 \cdot 10^8$m and $n = 10^8$ stars within the volume $V = \left(10^6 ly\right)^3$, we obtain

$$R_{\max}^{\text{Olb}} = 1.7 \cdot 10^{39}\text{m} = 1.7 \cdot 10^{23}\text{ly}.$$

This is 13 orders of magnitude larger than the value obtained from the Standard model.

The brightness of all stars, observed on earth, is therefore in the Standard Model smaller by the facto 10^{26}!

Furthermore, because of the expansion of the universe, the frequency of the light from stars is Doppler-shifted and the energy of the photons is smaller by the factor

$$h \cdot \nu / h \cdot \nu_0 = (1 - v/c)$$

which further decreases the observed brightness.

Note: In a static cosmos, the assumption of absorption of the star light by dark clouds would not solve the problem, because the absorbing matter would heat up and finally would have the same temperature as the stars, as soon as thermodynamic equilibrium has been reached. This means: **There will be no net absorption**.

6.1.5 Homogenity of the Universe

The fact, that all galaxies move away from the earth seems at a first glance to distinguish the earth, because it seems to be the center of the expansion. However, this is not the case as the simple consideration proves: We regard 3 Galaxies A, B and C (Fig. 6.9) with the distances $r_{AB}, r_{AC} and r_{BC}$. For the observer on A, the galaxies B and C recede from A, according to the Hubble law, with the velocities

$$v_{AB} = H_0 \cdot r_{AB}; v_{AC} = H_0 \cdot r_{AC}.$$

With the relation $r_{BC} = r_{AC} - r_{AB}$, we get

$$v_{BC} = v_{AC} - v_{AB} = H_0(r_{AC} - r_{AB}) = H_0 \cdot r_{BC}$$

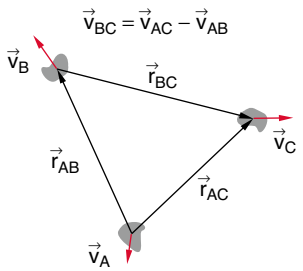


Fig. 6.9 Isotropy of the escape velocity of galaxies

.The observer B, therefore, sees the same isotropic expansion as the observer A.

Somehow more difficult is the answer to the question: From where the expansion of the universe has started? If it started from a certain point, this would distinguish this point and therefore violate the cosmological principle.

The answer to this question is possible, if we introduce a curved space, which can be exactly described within the General Relativity Theory. Its mathematical framework exceeds, however, the level of this introduction [12]. We restrict the representation therefore to a vivid more qualitative description.

6.2 The Metric of a Curved Space

We start with a two-dimensional world on the surface of a balloon onto which small slabs are glued, which represent the galaxies.

When inflating the balloon, the radius R of the balloon will increase and the model galaxies recede from each other. The distances between the galaxies increase proportional to the radius R and the velocity of this removal is proportional to $\mathrm{d}R/\mathrm{d}t$.

$$v_{AB} = \mathrm{d}/\mathrm{dt}(r_{AB}) \propto \mathrm{d}R/\mathrm{d}t. \tag{6.9}$$

Extrapolating into the past toward $\mathrm{t}=0$, the radius R of the balloon approaches zero ($\mathrm{R}(\mathrm{t}\to 0)\to 0$). The essential fact is that this point $\mathrm{R}\to 0$ is not a point on the balloon surface, but a point in the three-dimensional space of the balloon volume. The distinguished point where the expansion started therefore does not belong to the two-dimensional space of the galaxies but to a space with a higher dimension (in this example to the three-dimensional space into which the balloon is embedded.

In order to describe this more quantitatively, one has to introduce coordinates and a metric.

The surface of a sphere with radius R is described by

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 &= R^2, \\ x_1 &= R \cdot \sin\vartheta \cdot \cos\varphi, \\ x_2 &= R \cdot \sin\vartheta \cdot \sin\varphi, \\ x_3 &= R \cdot \cos\vartheta. \end{aligned} \tag{6.10}$$

For the line element dl on the surface, the metric is

$$\mathrm{d}l^2 = R^2\left[(\mathrm{d}\vartheta)^2 + \sin^2\vartheta \cdot (\mathrm{d}\varphi)^2\right]. \tag{6.11}$$

This shows that the metric on the surface of a sphere is defined by the radius R which does not belong to this surface, but represents a third dimension, namely the distance between any point of the surface to the center of the sphere, which does not belong to the surface. R is called the **scale factor** of the metric. It defines the curvature of the surface and is therefore also called the **radius of curvature**. The distance d between two points (ϑ_1, φ_1) and (ϑ_2, φ_2) on the surface is proportional to the scale factor R.

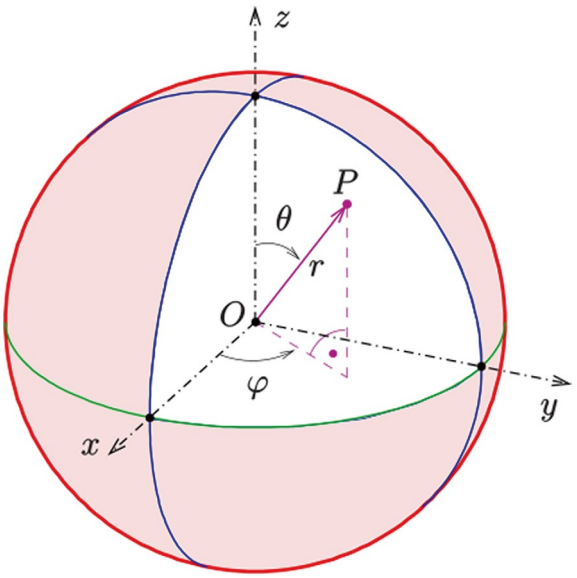


Fig. 6.10 Spherical coordinates

If we now proceed to a curved three-dimensional space (which is not accessible to a vivid perception), we have to introduce a fourth coordinate x_4.The three-dimensional surface of this four-dimensional sphere with scale factor R is described similar to (6.10) by

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 + x_4^2 &= R^2, \\ x_1 &= R \cdot \sin\psi \sin\vartheta \cos\varphi, \\ x_2 &= R \cdot \sin\psi \sin\vartheta \sin\varphi, \\ x_3 &= R \cdot \sin\psi \cos\vartheta, \\ x_4 &= R \cdot \cos\psi. \end{aligned} \tag{6.12}$$

This three-dimensional surface corresponds to the volume of the three-dimensional curved space. With the line elements

$$\begin{aligned} \mathrm{d}l_\psi &= R \cdot \mathrm{d}\psi; \\ \mathrm{d}l_\vartheta &= R \cdot \sin\psi\, \mathrm{d}\vartheta; \\ \mathrm{d}l_\varphi &= R \cdot \sin\psi \sin\vartheta\, \mathrm{d}\varphi \end{aligned} \tag{6.13}$$

we obtain the "volume" of this three-dimensional surface (i.e. the volume of the three-dimensional curved space)

$$V = \int_{\psi=0}^{\pi}\int_{\vartheta=0}^{\pi}\int_{\varphi=0}^{2\pi} R^3 \sin^2\psi \sin\vartheta \, d\psi \, d\vartheta \, d\varphi = 2\pi^2 R^3, \tag{6.14}$$

which is different from the two-dimensional surface $4\pi R^2$ of a sphere as well as from the three-dimensional volume $\frac{4}{3}\pi R^3$ of a sphere in the Euclidean space.

For the line element dl in our curved four-dimensional space, we obtain

$$dl^2 = dl_\psi^2 + dl_\vartheta^2 + dl_\varphi^2$$

$$= R^2\left[d\psi^2 + \sin^2\psi\left(d\vartheta^2 + \sin^2\vartheta \, d\varphi^2\right)\right]. \tag{6.15}$$

Introducing the three-dimensional distance $r = R \cdot \sin\psi$, the condition r = const. describes a sphere with radius r and surface $4\pi r^2$, as can be concluded from (6.12). The quantity r may be, for instance, the distance between the earth and a galaxy which is proportional to the scale factor R. When the scale factor R is changing during the expansion of the universe, the distance r also changes. We can write:

$$r = (R/R_0) \cdot r_0, \tag{6.16}$$

where we denote R_0 as the present scale factor at the present time t_0 and r_0 as the distance at $t = t_0$.

With the $r = R \cdot \sin\psi \Rightarrow dr = R \cdot \cos\psi \, d\psi$, we can eliminate the coordinate ψ using

$$\cos\psi = \left(1 - \sin^2\psi\right)^{1/2} = \left[1 - (r/R)^2\right]^{1/2} \tag{6.17}$$

and obtain for the line element in the curved three-dimensional space

$$dl^2 = \frac{dr^2}{1 - r^2/R^2} + r^2\left[(d\vartheta)^2 + \sin^2\vartheta \, (d\varphi)^2\right]. \tag{6.18}$$

It differs from (6.11) by the first term which depends on the scale factor R. The Eq. (6.18) describes the geometry of the curved three-dimensional space with a positive scale factor R, without using the fourth coordinate ψ.

The square root of dl^2 gives the distance between two neighboring points in our curved three-dimensional space. When the scale factor R increses, also the distance dl increases without a change of dr.

This is an Essential Understanding!.

▶ The expansion of our universe is really the increase of the scale factor R which is often called "**radius of curvature**" with reference to the two-dimensional curved surface of a sphere with radius of curvature R.

The limiting case $R \to \infty$

$$lim\,(dl)^2 = dr^2 + r^2\left[(d\vartheta)^2 + \sin^2\vartheta \, (d\varphi)^2\right], \quad R \to \infty \tag{6.19}$$

corresponds to the usual line element in the Euclidean flat three-dimensional space with $R = \infty$.

The factor $\kappa = 1/R^2$ describes the curvature of space. For $\kappa = 0$, we have a flat, i.e. not curved universe (Fig. 6.10b), for $\kappa > 0$, a spherical or elliptical universe (Fig. 6.10a) and for $\kappa < 0$, a hyperbolic universe (Fig. 6.10c) In Fig. 6.11, these three examples are illustrated for two-dimensional surfaces.

At the beginning of our universe, the scale factor R was very small. (**Note** that R has the dimension of a length).

Therefore, also the volume of the fireball $V = 2\pi^2 R^3$ was very small. Since $r = R \cdot \sin\varphi$ the distance r must be smaller than R. The spatial distance between two points in the universe

$$lim\,(dl)^2 = r^2\left[(d\vartheta)^2 + \sin^2\vartheta \, (d\varphi)^2\right] = 0, \quad R \to 0 \tag{6.20}$$

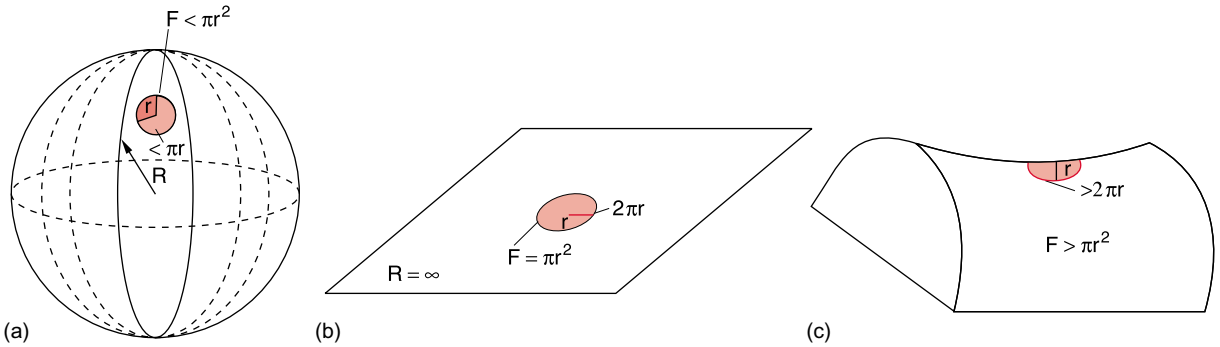


Fig. 6.11 Two-dimensional examples for the three-dimensional space with **a** R > 0 (spherical space); **b** R = ∞ (flat space); **c** R < 0 (hyperbolic space)

approaches zero for $R \to 0$. The important point is, however:

- **For $R \to 0$, the distance r tends for all points of the universe toward zero. No point is singled out! One cannot say that the big bang has happened at a certain point of the three-dimensional space. Remember the two-dimensional example of the inflatable balloon, where the point $R \to 0$ is not a point on the balloon surface.**

A further fact is remarkable: Although the volume $V = 2\pi^2 R^3$ is finite, it has no boundary similar to the curved surface of a sphere in the two-dimensional example.

When astronomers speak about the "*edge of the universe*", they just mean the farthermost distant galaxies. In our example of the surface of a sphere, these are the antipodes.

Such galaxies at the "*edge of the universe*" can be observed from any point of the universe and the distance to them is for each observer the same, namely about 10–15 Billion light years (see Sect. 6.3.10).

When we introduce a four-dimensional space–time metric by the relation

$$ds^2 = c^2 dt^2 - dl^2, \tag{6.21}$$

where c is the velocity of light, the world line of a photon is given by $ds^2 = 0 \to (dl/dt)^2 = c^2$.

If the photons are emitted at $t_0 < t_1$ from a source with the distance $r(t_0)$ from the observer, this distance has increased to

$$r(t_1) = [R(t_1)/R(t_0)] \cdot r(t_0) \tag{6.22}$$

at the arrival time t_1 at the observer. The photons have traveled a shorter distance than the true distance $r(t_1)$ of the galaxy at the observation time t_1 because we see the galaxy as it had been at the time t_0 of the photon emission (see Problem 6.5).

6.3 The Standard Model

In this section, the most important basic assumptions of the standard model of cosmology will be discussed and their comparison with experimental data outlined. Recommendable literature to this subject can be found in [6–9].

6.3.1 Radiation Dominated and Matter Dominated Universe

The cosmic background radiation can be described by a thermal Planck radiation at the temperature $T = 2.7\text{K}$. Its energy density (J/m^3) is (see Vol. 2, Eq. (12.28)).

$$w(T) = a \cdot T^4 \text{ with } a = \frac{4\pi^5 k^4}{15 h^3 c^3}. \tag{6.23}$$

For $T = 2.7\,\text{K}$, we get the numerical value

$$w(2.7\text{K}) = 4 \cdot 10^{-14} \text{Ws/m}^3.$$

This corresponds to an equivalent mass density

$$\varrho = w(T)/c^2 = 4.5 \cdot 10^{-31} \text{kg/m}^3. \tag{6.24}$$

This is smaller by several orders of magnitude than the present mass density of the universe, which is, according to presently acknowledged theories.

$$\varrho(t_0) \geq 4 \cdot 10^{-28}\ \text{kg/m}^3.$$

Therefore, we live today in a mass-dominated universe.

Due to the expansion of the universe, the mass density is decreasing as

$$\varrho(R) \propto 1/R^3, \tag{6.25a}$$

while the radiation density decreases faster as

$$w(R) \propto 1/R^4. \tag{6.25b}$$

The reason for this faster decrease is the frequency shift of photons during the expansion of the universe. If the universe expands with the velocity $v = dR/dt$, the photon energy decreases, because of the Doppler shift to

$$\begin{aligned} h \cdot \nu &= h \cdot \nu_0 (1 - v/c) \\ \Rightarrow \frac{\Delta\nu}{\nu} &= \frac{v}{c} = \left(\frac{1}{c}\right) \cdot \frac{dR}{dt} \end{aligned} \tag{6.26}$$

According to Hubble's law $dR/dt = H \cdot R$, the expansion velocity is proportional to the distance R, the energy loss of a photon $h \cdot \Delta\nu \propto R$. Since the density of photons decreases with $1/R^3$, the total energy loss of the radiation is proportional to $1/R^4$.

Extrapolating the mass density $\varrho(r)$ and the radiation energy density w(r) into the past, i.e. $R \to 0$, we get the curves shown in Fig. 6.12, which intersect at a time $t_s \approx 300.000$ years after the big bang. This shows: For $t < t_s$

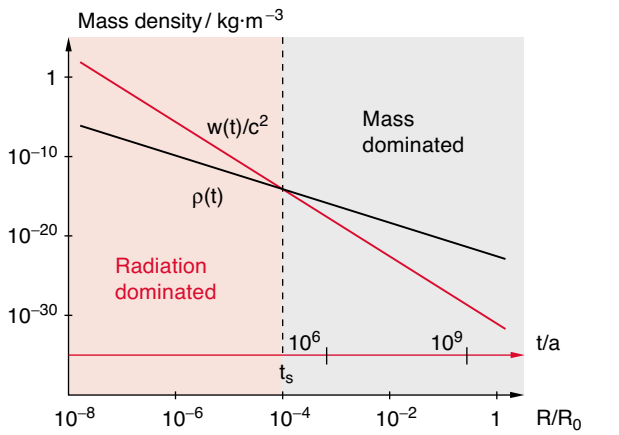


Fig. 6.12 Mass density ρ (black curve and radiative energy density (red curve) as a function of the scale parameter R(t) of the expanding universe

a radiation dominated universe existed, while for $t > t_s$ the matter dominated. Since $w(T) \propto T^4$ and $w(R) \propto R^{-4}$, it follows that

$$T(R) \propto 1/R. \tag{6.27}$$

Therefore, we get from the radiation temperature $T(t_0) = 2.7\text{K}$ measured today in the universe with radius R_0 the temperature at a time t where the radius was R, a radiation temperature

$$T(R) = (R_0/R) \cdot 2.7\text{K}. \tag{6.28}$$

Extrapolating from the present scale factor $R_0 \approx 14 \cdot 10^9$ly to an earlier time of the universe, where $R = 10^9$m (diameter of the sun), we get the incredibly high temperature of

$$T(R) = 10^{17}\ \text{K}$$

and the radiation energy density

$$w(T) = 10^{55}\ \text{J/m}^3.$$

This corresponds to the equivalent mass density of

$$\varrho(T) = w(T)/c^2 = 10^{38}\ \text{kg/m}^3.$$

This volume with a scale factor $R < 10^9$m in the early phase of the universe, filled with an extremely high energy density at a temperature $T > 100$ Trillion Kelvin, is called **the fire ball.**

At these high temperatures, the repulsive expansion pressure is much larger than the attractive gravitational pressure. The fire ball therefore expands at a high radial velocity while it cools down. This is called the **big bang**.

6.3.2 Hubble Parameter and Critical Density

We regard an arbitrary galaxy with mass m, which has the distance r from the earth and recedes with the velocity $v(r) = H \cdot r$ (Fig. 6.13). Its kinetic energy is

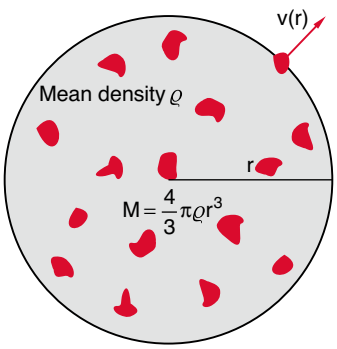


Fig. 6.13 Illustration of the deceleration of the escape velocity v® of galaxies by the gravitational force exerted by all galaxies inside the volume with radius r

$$E_{\text{kin}} = \frac{1}{2}mv^2 = \frac{1}{2}mH_0^2r^2. \tag{6.29}$$

Its potential energy is determined by the gravitational force caused by all galaxies inside the sphere with radius r. We, therefore, get

$$E_{\text{pot}} = -G \cdot \frac{m \cdot M(r)}{r}, \tag{6.30}$$

where $M(r) = \frac{4}{3}\pi r^3\overline{\varrho}$ is the total mass of all galaxies within the sphere with R ≤ r and $\overline{\rho}$ is the mean mass density in this sphere. Inserting this into (6.30) gives the total energy

$$E = E_{\text{kin}} + E_{\text{pot}} = mr^2\left(\frac{1}{2}H_0^2 - \frac{4}{3}\pi G\overline{\varrho}\right). \tag{6.31}$$

For the critical mass density

$$\varrho_c = \frac{3H_0^2}{8\pi G}. \tag{6.32}$$

The total energy is E = 0. In this case, the expansion velocity v ∝ 1/r tends for $r \to \infty$ toward zero. This implies that also the time derivative $dR(t)/dt$ of the scale factor R approaches zero for $r \to \infty$. The final fate of the universe is a static equilibrium state where no processes any longer occur.

For $\overline{\varrho} < \varrho_c$ the total energy remains positive, i.e. the expansion will not be stopped. This scenario is called an **open universe**.

For $\overline{\varrho} > \varrho_c$ the universe is closed, the expansion stops when the kinetic energy is completely transferred into potential energy, and then starts to collapse. At the end of this collapse, probably a singularity (Fireball) would happen again and an oscillating motion of the universe would be the result. In Fig. 6.13, the three different possibilities are illustrated.

Presently, we do not know the numerical value of the critical density ϱ_c sufficiently accurate in order to determine which of the three possibilities are realized in our universe. Measurements indicate so far that the case $\overline{\varrho} > \varrho_c$ might be realized.

More recent accurate measurements of the distances and velocities of supernovae Ia seem to confirm that the expansion velocity increases. The reason for this astonishing result is still a matter of controverse discussions. A mysterious "**dark energy**", which would cause a repulsive force, might be a solution. But up to now, nobody can give a physical reasonable explanation about the dark energy (see Sect. 6.6.2).

Since the expansion causes a decrease of the density ρ, the Hubble parameter must have been larger in former

times. If we assume that the present density is close to the critical density, we can write according to (6.32)

$$H(t) = \sqrt{\frac{8\pi G}{3}\varrho(t)}. \tag{6.33}$$

Remark The description of the expansion of the universe should use the scale parameter R(t) instead of the geometric distance r(t). However, because of the relation.

$$r(t) = (R/R_0) \cdot r_0, \tag{6.34}$$

i.e. $r \propto R$

the result of (6.32) does not change when using r instead of R.

For the experimental determination of the present density $\rho_{0,}$ the following possibilities can be used:

- At first the critical density ϱ_c must be determined. According to (6.32), it depends on the Hubble constant H_0. The value of H_0 can be obtained in two different ways:

(1) From the relation $v = H_0 \cdot r$, the Hubble constant is determined if the expansion velocity and the distance of a galaxy has been measured. The velocity can be obtained from the Doppler shift (red-shift) of spectral lines emitted from the galaxy. However, some perturbing effects can falsify the result, for instance, the motion of our own galaxy, or the redshift of light from far distant galaxies, caused by scattering and absorption when the light passes through dust clouds, because the scattering cross section $\sigma \propto \lambda^{-4}$ strongly depends on the wavelength λ. The blue part of a spectral line is more scattered than the red part. This results in a red-shift.

(2) The Hubble constant can be also obtained from the estimated age τ of the universe, which is related to the Hubble constant by $H_0 = \frac{2}{3}/\tau$ as will be shown in Sect. 6.3.8. Based on spectral measurements of spectra of the oldest stars in globular clusters and on star models, the age of these stars can be estimated as $T \geq 0.8 \cdot 10^{10}$a. Since the age of the universe must be larger than the early stars, it follows: $\tau > T$. This gives $H_0 < 83\ \mathrm{kms}^{-1}\,\mathrm{Mpc}^{-1}$.

(3) The distance r can be nowadays accurately measured, if a supernova Ia is present in the galaxy, which can serve as a calibrated distance candle (see Sect. 6.5.10).

The presently accepted optimum value of $H_{0,}$ obtained from the measurements discussed above, is

$H_0 \approx 70.5\ \mathrm{kms}^{-1} \cdot \mathrm{Mpc}^{-1}$.

This gives the critical density $\varrho_c = 0.93 \cdot 10^{-26}\mathrm{kg/m}^3$, which corresponds to about 5 H-atoms per m^3.

- The mean mass density $\overline{\varrho}$ in the universe can be estimated by determining the mass of a galaxy from its luminosity and its rotation angular velocity. Multiplication with the total number of galaxies in the universe gives the total mass (see Sect. 6.5). The mean density, obtained in this way is about $\varrho = 0.06\varrho_c$. This value is too small because the luminosity measurements do not gather the dark matter.
- Another method of estimating the total mass of all galaxies from the measured velocity dispersion of galaxies in a galaxy cluster, using the Virial theorem. This gives mean densities $\overline{\rho} \approx 0.2\varrho_c$. similar results are obtained from the determination of the space curvature around galaxies, which depends on the mass of the galaxy.
- Another way to determine the mean density $\overline{\varrho}$ are measurememts of the isotope ratio N(D)/N(H) of the numbers

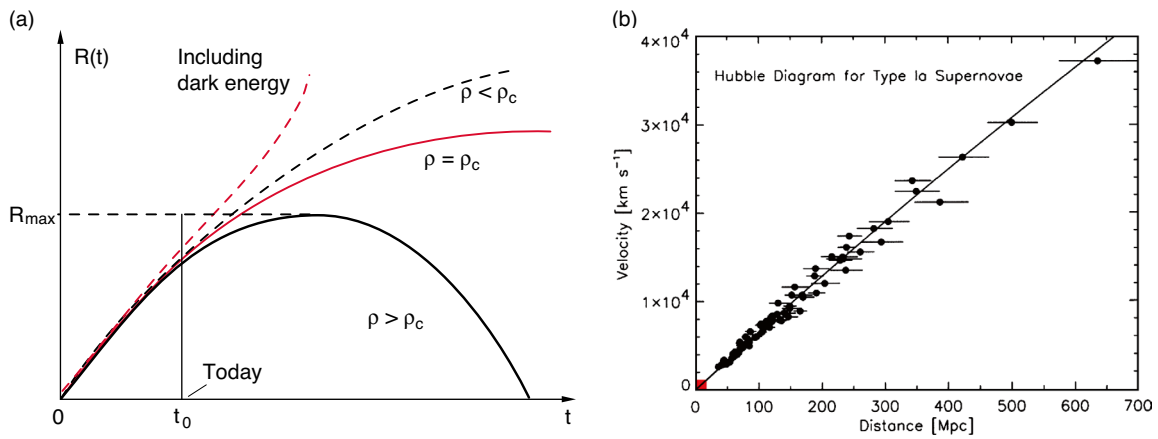


Fig. 6.14 **a** Time dependence of the scale parameter R(t) for different mean mass densities $\overline{\rho}$. **b** Experimental points with error bars of the expansion velocity of galaxies versus their distance (Wikipedia)

of deuterium and hydrogen. Detailed calculations within the framework of the Standard Model have determined the theoretically expected ratio during the formation- and decay phase of deuterium in the Fireball at temperatures of 10^9 K as the function of the baryon density at that time (see Sect. 6.3.4). The ratio

$$N\left({}^2_1\mathrm{H}\right)/N\left({}^1_1\mathrm{H}\right) = 2 \cdot 10^{-5}$$

measured today is in accord with the calculations only, if the baryon density at that time had been $\varrho_\mathrm{B}\left(T = 10^9\,\mathrm{K}\right) = 2 \cdot 10^2\ \mathrm{kg/m^3}$ (Fig. 6.15).

Since the total number of baryons has been constant during the further expansion of the universe, the density today $\varrho(T = 2.7\mathrm{K})$ must be equal to the product of the density at that time $\varrho_\mathrm{B}\left(T = 10^9\mathrm{K}\right)$ times the ratio of the volumes of the universe

$$\varrho_\mathrm{B}(2.7\mathrm{K}) = \varrho_\mathrm{B}\left(10^9\mathrm{K}\right) \cdot (R_0/R)^3$$

$$= \varrho_\mathrm{B}\left(T = 10^9\mathrm{K}\right) \cdot \left(2.7/10^9\right)^{-3} \qquad (6.35)$$

because the temperature $\mathrm{T} \propto 1/\mathrm{R}$ is inversely proportional to the scale factor R.

- If the expected density in the universe today is calculated from (6.35) the value $\varrho = 4 \cdot 10^{-28}\mathrm{kg/m^3}$ is obtained which is about 5% of the critical density.
- According to these results, our universe should be open. There are, however, several possibilities of additional mass, which increase the mean density.
- The intergalactic matter may contain as much or even more mass than the galaxies. Furthermore, cold celestial objects, such as brown or black dwarfs or black holes, which do not contribute to the luminosity, can add mass to the total mass in the universe. They are not considered when estimating the mass from the luminosities of galaxies.

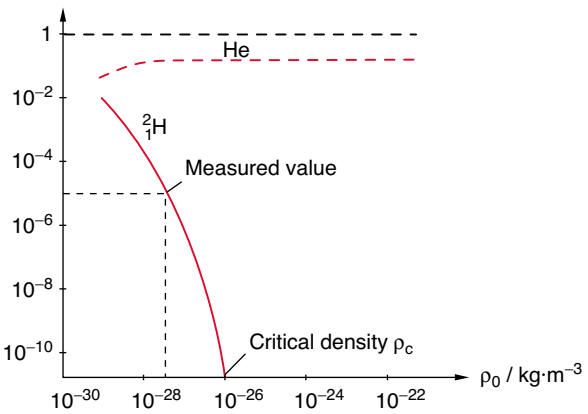


Fig. 6.15 Mass fraction of Deuterium and Helium as a function of the present mass density ρ_0 (after calculations of Wagner: "Zur Synthese leichter Kerne im Feuerball"

- Since neutrinos have a finite rest mass (the lower limit is today for the electron neutrinos $m(\nu_\mathrm{e}) \cdot c^2 \leq 10\mathrm{eV}$, for the μ on neutrinos $m\left(\nu_\mu\right) \leq 170\mathrm{keV}$ and for the τ-neutrinos $m(\nu_\tau) \leq 18\mathrm{MeV}$), they can contribute, because of the large number in the universe, a non-negligible contribution to the total mass.
- The "dark matter" (see Sect. 6.6.1) provides probably the largest contribution to the mass density.

More recent estimations of the intergalactic matter and measurements of the neutrino masses confirm that the first two contributions do not essentially add to the mass density, which leaves the dark matter as the most probable explanation for the missing mass, although it is still not known, what the dark matter really is, which particles hide behind this expression. We will discuss this point in more detail in Sect. 6.6. The relative contributions of the different forms of energy resp, their mass equivalents are illustrated in Fig. 6.16.

Up to now, it cannot be definitively decided whether the mass density is smaller or larger than the critical density, i.e. whether our universe is open or closed.

Due to theoretical and philosophical arguments, the value $\varrho = \varrho_\mathrm{c}$ is preferred.

The measurements of recession velocities of galaxies meet the following difficulties which can distort the results:

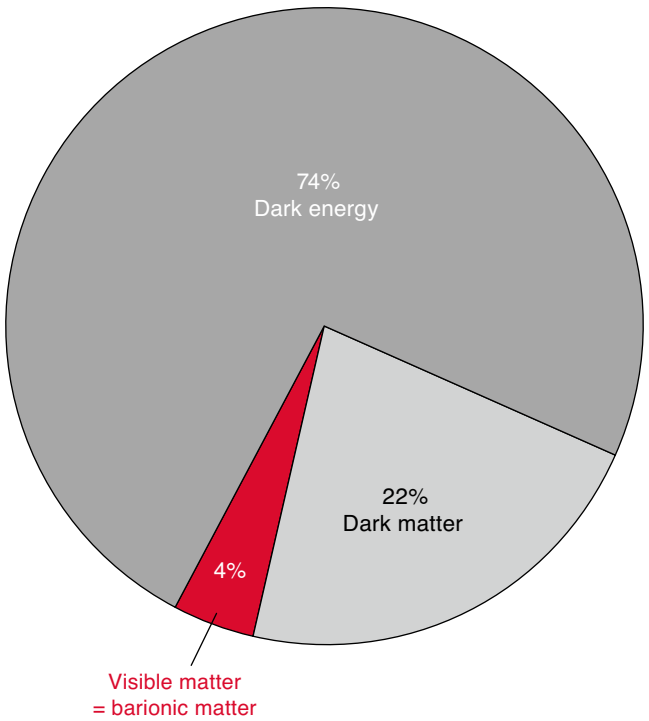


Fig. 6.16 Contribution of baryonic matter (red sector), dark matter (gray sector) and dark energy (black sector) to the total energy density in the universe

(1) If a galaxy is a member of a galaxy cluster, it is affected by the gravitational attraction by all other members of the cluster. Therefore, the general recession velocity v of the cluster is superimposed by the so called peculiar velocity v_p of an invidual galaxy. One has to determine the magnitude and direction of v_p in order to extract from the measured redshift the contribution of the recession velocity v.
(2) If the light from a galaxy passes through clouds of gas and dust, the blue part is more absorbed and scattered as the red part, thus causing a redshift that is not due to the recession velocity.
(3) One has, of course, to take into account the velocity of the earth, which is not only due to its revolution around the sun, but also to the rotation of our galaxy and its motion due to the gravitational attraction by neighboring galaxies.

6.3.3 The Early Phase of the Universe

Using known physical laws the development of the universe can be traced back up to fractions of a second after the big bang. About the time before, only speculations and not proved assumptions can be found, because up to now, no unified theory of quantum theory and gravitation theory exists. The behavior of matter = energy under the extreme conditions at the very early time ($t < 10^{-44}$ s) after the big bang) is widely unknown. The earliest time after the big bang where we can assume that known laws of physics can be applied, is called the Planck time

$$t_{Pl} = \left(\hbar \cdot G/c^5\right)^{1/2} = 5.4 \cdot 10^{-44}\text{s}. \tag{6.36a}$$

This is the time the light traverses the Plank length

$$L_{Pl} = \left(\hbar \cdot G/c^3\right)^{1/2} = 1.6 \cdot 10^{-35}\text{m}. \tag{6.36b}$$

Theoretical models assume that the time before t_{Pl} does not proceed continuously, but in small steps. Therefore, one would need a quantum gravitation theory for the proper description of times smaller than t_{Pl}. It can be shown that each object that passes through a time span smaller than the Planck time would become a singularity in the classical description.

We therefore assume that the big bang started with a singularity $\varrho(t = 0) = \infty$. In order to obtain the expansion times for the scale parameter we start with the Hubble parameter

$$H(t) = \sqrt{\frac{8\pi G \cdot \varrho(t)}{3}}. \tag{6.37}$$

For the density $\varrho(t)$, we obtain

$$\varrho(t) \propto [1/R(t)]^n \tag{6.38}$$

with n = 3 for the mass dominated universe and n = 4 for the radiation dominated universe (Fig. 6.11). Then it follows for the Hubble parameter

$$H(t) \propto [1/R(t)]^{n/2}. \tag{6.39}$$

The expansion velocity is then

$$v(t) = \mathrm{dR/dt} = H(t) \cdot R(t) \propto [R(t)]^{1-n/2}, \tag{6.40}$$

from this we can conclude

$$\int \mathrm{d}t = t_2 - t_1 \propto \int_{R_1}^{R_2} \frac{\mathrm{d}R}{R^{1-n/2}}$$

$$= \frac{2}{n}\left[\frac{1}{H(R_2)} - \frac{1}{H(R_1)}\right]. \tag{6.41}$$

Inserting $t_1 = 0 \Rightarrow R(t_1) = 0$, we get with (6.41) for the time between the big bang and the expansion to the scale parameter R(t) at the time t

$$t = \frac{2}{n}\sqrt{\frac{3}{8\pi G\varrho(t)}} \propto R^{n/2}, \varrho \propto \frac{1}{R^n}. \tag{6.42}$$

Inserting the expression (6.37) for H, we get from (6.41)

$$t_2 - t_1 = \frac{2}{n}\sqrt{\frac{3}{8\pi G}}\left[\frac{1}{\sqrt{\varrho(t_1)}} - \frac{1}{\sqrt{\varrho(t_2)}}\right]. \tag{6.43}$$

For the radiation-dominated age, ($t < t_s$ in Fig. 6.11) is n = 4. The expansion time is then $t \propto R^2$ or $R \propto \sqrt{t}$.

For the mass-dominated universe, ($t > t_s$) is $n = 3 \Rightarrow R(t)t^{2/3}$.

Note:

- The scale parameter R(t) does not increase linearly with time, It is:
  $R(t) \propto \sqrt{t}$ for the radiation-dominated universe.
  $R(t) \propto t^{2/3}$ for the mass-dominated universe.

Remark We have calculated in Sect. 5.2 x and y the collapse time of objects (gas clouds or stars) from the radius R_0 down to R = 0

$$t_c = \sqrt{\frac{3\pi}{16G \cdot \varrho}}. \tag{6.44}$$

Based on symmetry reasons, we expect a similar time for the expansion from R = 0 to R = R_0. In fact, Eq. (6.42) is for n = 3 (mass-dominated universe) with $t = \pm 1/(6\pi G\varrho)$ nearly (aside from a factor 3) identical to (6.44).

In Fig. 6.17, a timetable of the early development of the universe is compiled, according to the Standard model. From the big bang at t = 0 until the Planck time $t = 10^{-43}$ s, the physics of the universe is undefined, because the physical laws, as we know them today, cannot safely extrapolated into the time of extreme conditions of temperature and pressure All four interactions (strong, weak el. Magn. And gravitational interactions) are unified, which means undistinguishable. At $t = 10^{-43}$ s a phase transition occurs where the gravitation is separated from the other 3 interactions. The temperature is about $T = 10^{32}$ K and the universe consists of a quark- gluon plasma, where no distinction between the different elementary particles is possible. This development stage is called the GUT-era (era of Grand Unification Theory), because the unification of the three remaining interactions is described by this theory.

At the time $t = 10^{-35}$s ($T = 10^{28}$ K), a second phase transition takes place when the strong interaction separates from the electro-weak interaction. Now hadrons and leptons are no longer equivalent but can be distinguished. This phase transition releases energy. Now the inflation period starts according to the theory of *A. Guth* [13] which lasts until $t = 10^{-33}$ s.

At $t = 10^{-11}$ s ($T = 10^{15}$ K), the electro-weak symmetry breaks up and the leptons separate into distinguishable electrons, neutrinos and their antiparticles.

At $t = 10^{-5}$ s ($T = 10^{12}$ K), the Quark-Hadron symmetry breaks up and protons and neutrons are formed. During all these processes, the universe expands further and cools down. As long as the temperature T is high enough (kT must be larger than the mass energy of the particles), thermal equilibrium exists and the rates of formation and destruction of particles are equal. This is the case for $T > 10^{12}$ K.

The relation between temperature and expansion time during the period of thermal equilibrium can be obtained as follows.

The energy density of the radiation is ϱT^4, it decreases due to the expansion $1/R^4$. Since ϱR^{-3} and $R\sqrt{t}$, it follows $Tt^{-1/2}$.

The temperature T decreases, therefore, with increasing time $T1/\sqrt{t}$.

This early stage of the universe $t < 10^{-15}$s and $T > 10^{17}$ K is called **Fireball**. According to the Standard Model, the "*hot primeval soup*" of the fireball at a time $t < 10^{-15}$ s consists of a quark-gluon plasma.

Time era	Time in sec	Temperature in K	State
	$t = 0$	?	Big bang
Planck era			All 4 interactions are unified, i.e. not distinguishable
	$t = 10^{-43}$	$T = 10^{32}$	
GUT-era			Gravitation decoupled
	$t = 10^{-35}$	$T = 10^{28}$	The other 3 interactions are still unified
			(Grand Unification), quark-gluon plasma
Inflation phase			Strong and weak interactions are separated
			Universe expands by a factor q = 10^{40}–10^{50}
	$t = 10^{-32}$	$T = 10^{25}$	
Quark-era			The quark-gluon plasma consists of quarks, leptons and photons
	$t = 10^{-12}$	$T = 10^{14}$	
	$t = 10^{-7}$		
Hadron-era			Protons and neutrons and their anti-particles are formed; Neutrinos decouple
	$t = 10^{-4}$	$T = 10^{12}$	
Lepton-era			Electrons and positrons annihilate, Myons decay
	$t = 10$	$T = 10^{10}$	
		$T = 3 \cdot 10^9$	
Nucleon synthesis			He-, D- and Li-nuclei are formed
	$t = 3$ min	$T \approx 10^9$	

Fig. 6.17 Timetable of the first 3 minutes after the big bang

Later on ($t > 10^{-15}$ s the quark-gluon plasma converts into a dense mixture of particles and photons. All particles and photons are at thermal equilibrium. The density is so high that even neutrinos are completely absorbed in spite of their small absorption- and scattering cross section.

At a later time ($t = 10^{-4}$ s) with the temperature $T = 3 \cdot 10^{11}$ K, the scattering and absorption rate of neutrinos becomes so small, that they are no longer at thermal equilibrium with the other particles. This is called the decoupling of the neutrinos.

We can now look at different later stages of the early universe:

- At temperatures of $T = 10^{14}$K ($t = 10^{-8}$ s), the kinetic energy of the particles ($kT \approx 10^{10}$eV) is large compared with the rest mass energy of the hadrons ($m_p c^2 \approx 10^9$eV). The particles in the fireball can produce by collisions new particles or by collisions with their antiparticles annihilate into photons. By recombination of two photons, a particle–antiparticle pair can be generated. In this era of the fireball, equilibrium exists between the production and destruction of particles, i.e. the fireball is at thermal equilibrium. This stage where hadronic processes dominate is called the **era of hadrons**. The number of hadrons is equal to that of photons.

All particles and photons are at thermal equilibrium. The density is so high that even neutrinos are completely absorbed in spite of their small absorption- and scattering cross section.

The standard model shows, that the ratio of photons to baryons observed today allows the deduction of the ratio of particles to antiparticles at the time $t = 10^{-4}$s, which was $\left(1 + 10^{-9}\right)$. This means that there had been nearly the same number of particles and antiparticles with a slight $\left(10^{-9}!\right)$ excess of particles.

The cause of this slightly unbalanced number is the asymmetry of the production rates which is for particles larger by the factor $\left(1 + 10^{-9}\right)$.

While at sufficiently high temperatures thermal equilibrium exists, for decreasing temperatures, the exothermic reactions are preferred over the endothermic reactions. This means if energy is gained when a particle X is transferred into other particles, the decay rate of X will be larger than the production rate of X and the concentration of X continuously decreases.

- About 10^{-4}s after the big bang, the temperature had decreased to about 10^{12}K and the scale factor had increased to $R = 10^{14}$m $= 700$AU. Now no longer hadrons can be produced because the thermal energy ($< 10^8$eV) is now smaller than the mass energy of proton or neutron ($\approx 10^9$eV). The hadrons can still annihilate with their antiparticles and produce photons.

Because of the slight asymmetry between the number of hadrons and anti-hadrons, after the annihilation, a small rest of 10^{-9} of the former hadron number is leftover. This rest forms the basis of matter in our present world, while the by far larger share of matter (99.9999999%) has been annihilated to photons.

The slight asymmetry between hadrons and anti-hadrons is probably due to the violation of the CP-symmetry, which had been first discovered by *James Cronin* and *Val Fitsch* for the decay of the neutral Kaons $K^0 = (d\bar{s})$ and its antiparticle $K^0 = (d\bar{s})$ [14]. These Kaons decay via the weak interaction into two or three pions and the decay rate differs slightly for the anti-Kaon from that of the Kaon.

A further violation of the CP symmetry was found at CERN where collisions of protons against protons D^+–mesons $((c\bar{d}))$, $D^0((c\bar{u}))$ and $\overline{D}_0$ – mesons $(\bar{c}u)$ and D^- mesons $(c\bar{d})$ were detected, which decay into $\pi\pi$ - or KK-pairs, according to the reaction scheme

$$\begin{aligned} D^0 \quad &\rightarrow K^- + \pi^+; \overline{D}_0 \rightarrow K^+ + \pi^-, \\ D^+ &\rightarrow K^- + \pi^+ + \pi^+; D^- \rightarrow K^+ + \pi^- + \pi^-. \end{aligned} \tag{6.45}$$

For the anti-D-mesons $\overline{D}$ again a distinctively shorter decay time was measured than for the D^+-mesons.

A similar asymmetry was found for the decay of the B-mesons and their anti-mesons.

If for times $t < 10^{-4}s$ in the early universe, the decay times for the anti-baryons were shorter than for the baryons the number of baryons was slightly larger, and at the annihilation of baryon–antibaryon pairs a small rest of baryons survived.

The nucleons existing today as building blocks for atomic nuclei originate from the early stage of the universe with $t < 10^{-4}s$!

The annihilation of hadrons with anti-hadrons decreased the number of hadrons by a factor 10^9. Now the ratio of hadron density and photon density is only 10^{-9}. This ratio remains constant during the whole future development of the universe because both densities decrease during the expansion as $1/R(t)^3$.

Note that the photon number density n_{ph} decreases $\propto R^{-3}$ but the radiation energy density $n_{Ph} \cdot h\nu$ is $\propto R^{-4}$.

- Leptons can still be created and destroyed because of their smaller mass. Therefore, in the subsequent **age of leptons** (from $t_1 = 10^{-4}$s, $T_1 = 3.10^{11}$K up to $t_2 = 1$s, $T_2 < 10^{10}K$) (Fig. 6.18) they are at thermal equilibrium with the radiation field, according to the equilibrium equation

i.e. the annihilation rate equals the production rate.

$$e^- + e^+ \leftrightarrow h \cdot \nu, \tag{6.46}$$

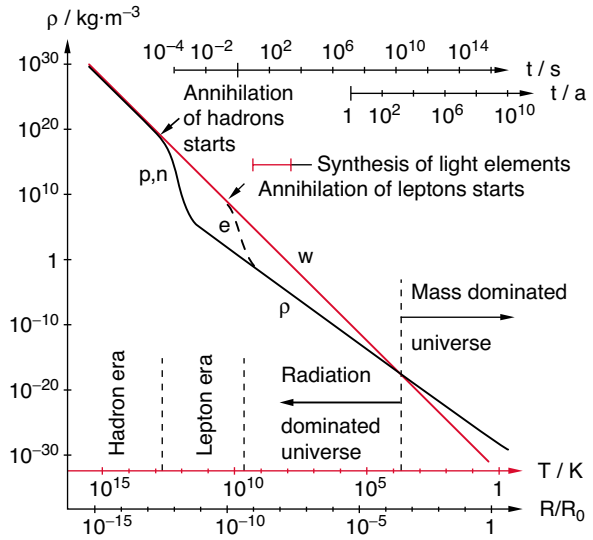


Fig. 6.18 Summarizing representation of the different time intervals during the development of the universe

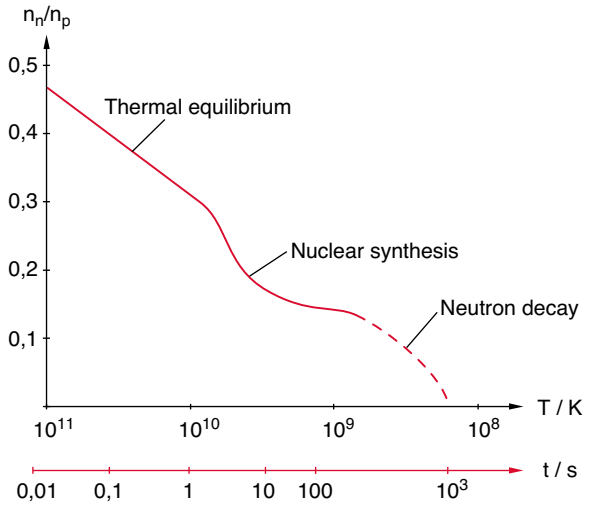


Fig. 6.19 Ratio N_n/N_p of the number of neutrons and protons as a function of temperature (after St. Weinberg: The first three minutes (DTV München 1980))

Also, nuclear reactions occur where leptons participate. For instance, the reactions

$$p + e^- \leftrightarrow n + \nu_e; \; n + e^+ \leftrightarrow p + \overline{\nu}_e \qquad (6.47)$$

convert neutrons into protons and vice versa.

The equilibrium ratio of neutrons and protons is determined by the temperature T and by the mass difference $\Delta m = m_n - m_p = 2.32 \cdot 10^{-30}$kg between neutrons and protons. With $\Delta E = \Delta m \cdot c^2$ the ratio is

$$\frac{m_n}{m_p} = e^{-\Delta E/kT}. \qquad (6.48)$$

Example

For $T = 10^{10}$ K $\Rightarrow$ KT $= 8.6 \cdot 10^6$eV, $\Delta E = 1.3 \cdot 10^6$eV $\Rightarrow m_n/m_p \approx e^{-1,5} \approx 0.22$. ◀

In Fig. 6.19, the ratio n_n/n_p is plotted as a function of the temperature and the development time of the universe.

Remark One might be surprised that the expansion of the universe from $R = 0$ to $R = 10^{14}$ m $(0.01ly)$ proceeds in a time span of 1 s. This corresponds to the expansion velocity $v = 10^{14}$ m/s, which is 10^6 times larger than the velocity of light. This does not contradict the special relativity theory, because this is the velocity dR/dt of space expansion not the velocity of particles.

6.3.4 The Synthesis of the Light Elements

In the following time span from 1 s to about 100 s, the temperature drops to $T < 10^9$K. The photon energy is now not high enough to generate electron–positron pairs ($2m_ec^2 = 1$MeV $\rightarrow$ kT $= 10^6$ eV $\rightarrow$ T $= 10^{10}$K). Therefore, the existing lepton–anti-lepton pairs annihilate and produce photons. However, since there is a small excess of $(1 + 10^{-9})$ electrons over the positrons, a small fraction of electrons remain. The number of leftover electrons is equal to the number of leftover protons. Therefore, the matter is electrically neutral. The electrons are still coupled by the Compton effect with the radiation field. This allows the energy exchange between photons and electrons.

The neutrinos are already decoupled from the radiation field. They fly freely with the velocity of light c through the universe.

Since there are no longer electrons with sufficient energy to produce neutrons according to the reaction

$$p + e^- + \overline{\nu}_e \rightarrow n \qquad (6.49)$$

new neutrons are no longer produced. The existing neutrons decay with their half lifetime of 887 s into protons

$$n \rightarrow p + e^- + \overline{\nu}_e. \qquad (6.50)$$

However, before all neutrons have disappeared, they can collide with protons and form deuterium nuclei

$$p + n \rightarrow {}^2_1\mathrm{H} + \gamma. \qquad (6.51)$$

Because of the small binding energy of deuterium nuclei ($E_B \approx 1$MeV), they will dissociate by collisions at temperatures above 10^9K. After a time $t > 10^2$s, when the temperature has dropped below 10^9K, a noticeable density of deuterium nuclei builds up (Fig. 6.20).

The relative deuterium concentration measured today is a critical test of the results of the standard model (Fig. 6.15).

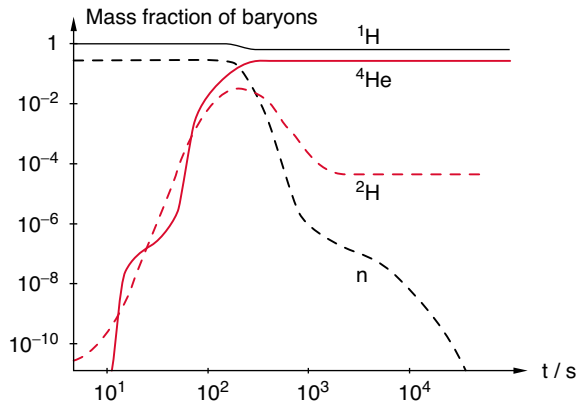


Fig. 6.20 Densities of H, D and He nuclei and neutrons during the synthesis in the hot early universe

After sufficient numbers of deuterium nuclei have been formed, the synthesis of Helium nuclei can start, following the reaction

$$^2_1\mathrm{H} + ^2_1\mathrm{H} \rightarrow ^4_2\mathrm{He}. \tag{6.52}$$

According to model calculations of the standard theory, helium synthesis starts about at times $t = 220\,\mathrm{s}$ after the big bang at t = 0, when the temperature has dropped down to about $0.9 \cdot 10^8$ K. With the proton number density n_p and the neutron number density $n_n < n_p$ at most $n_n/2$ deuterium nuclei can be formed, and the remaining number of protons is

$$n_p(t > t_2) = n_p(t_2) - n_n(t_2).$$

The ratio of helium atoms to hydrogen atoms is then

$$\frac{n(^4\mathrm{He})}{n(^1\mathrm{H})} = \frac{n_n(t_2)}{2 \cdot [n_p(t_2) - n_n(t_2)]}. \tag{6.53}$$

Even if at the time t_2 a new equilibrium state will be ensued at a lower temperature, the ratio n_n/n_p of (6.48) is no longer valid because the neutrons decay and new neutrons are no longer produced at the lower temperature. At the time t_2, the ratio n_n/n_p is about 0.14.

$$\frac{n(^4\mathrm{He})}{n(^1\mathrm{H})} = \frac{0.14}{2 \cdot (1 - 0.14)} = 0.082. \tag{6.54}$$

The relative mass amount of helium compared to the total mass m(H+D+He) is then about 25%. This agrees with the value observed today (Fig. 6.20).

The free neutrons are nearly completely used for the synthesis of ^{4}He. A small rest remains to build up ^{3}He and ^{7}Li.

After three minutes, the temperature drops to an extent that the synthesis of heavier nucleons is not possible. All elements above Li are formed much later by fusion processes in the central part of stars (see Sect. 6.5x), The heavy elements beyond iron are produced in the hot star envelopes during supernova explosions (see Sect. 6.5y). The first three minutes in the development of the universe are therefore essential for the production of the five lightest elements.

6.3.5 Transition from the Radiation-Dominated to the Mass-Dominated Universe

About 100 s after the big bang, the temperature has dropped below 10^9 K. This corresponds to a mean energy of the radiation field of 10^5 eV-. The energy of the photons is no longer high enough to produce electron–positron pairs, but the pairs of particle and antiparticle can still annihilate at collisions of electrons with positrons to produce photons. Since their density decreases with R^{-3} such annihilation processes become more and more seldom. Since the mass density decreases with $1/R^3$ but the radiation energy density with $1/R^4$, there is a time t_c when the mass energy density equals the radiation energy density. For $t > t_c$, the universe becomes mass-dominated (Fig. 6.12). This time t_s is about 48.000 years after the big bang when the temperature has dropped to $T = 10^4$K.

Example

With the radiation density $w(2.725\mathrm{K}) = 4 \cdot 10^{-14}\mathrm{J/m^3}$ observed today, the extrapolation back to $T = 10^9$K $(R = 5 \cdot 10^{-8}).R_0(t_0)$ gives the radiation energy density $w(10^9\mathrm{K}) = 7 \cdot 10^{20}\mathrm{J/m^3}$, while the mass density is $\varrho(t = 10s; T = 10^9\mathrm{K}) = 0.1\ \mathrm{kg/m^3}$, which corresponds to a mass energy density of $\varrho c^2 = 10^{16}\mathrm{J/m^3}$. At this temperature of 10^9K, and at the time t = 10–100 s, the universe is still radiation-dominated. Only at temperatures $T \approx 10^4 K$ (t = 48.000 years after the Big Bang), the transition to the mass-dominated universe occurs. For $T = 10^4 K$ is $w = 7\mathrm{J/m^3}$ and $\varrho c^2 = 10\mathrm{J/m^3}$. ◀

Radiation and matter are still coupled with each other because the atoms are still ionized and the free electrons scatter the photons sufficiently strong to prevent them from escaping, i.e. the photons are captured within the matter-filled space.

6.3.6 Decoupling of Matter and Radiation

During the next 340.000 years, the universe further expands and cools down. When the temperature of the expanding

universe has decreased to $T = 3000$K (this is the case at about 80.000 years after the big bang), the electrons can recombine with the protons to form neutral hydrogen atoms. Sometime earlier (at $T = 8000$K), helium nuclei can recombine with electrons to form neutral helium atoms. However, since the density of He-atoms is much smaller than that of hydrogen atoms, this has only a minor effect.

The recombination of electrons with ions has a fundamental effect: The continuum radiation can no longer interact with free electrons, which causes a nearly complete vanishing of the radiation pressure. While the photons before this recombination were scattered by the free electrons and their path through matter consists of statistically varying pieces with the mean free path length Λ and statistically distributed directions (Fig. 6.21) they can now freely move on straight lines through the universe.

This expanding and cooling down universe was not completely uniform but showed local density fluctuations. There has been a long-lasting discussion about how these local density fluctuations have emerged and whether they are responsible for the relatively early emerging structures in the universe that have led to the formation of galaxies and stars.

Since before the decoupling of matter and radiation, the matter density fluctuations must have caused similar local energy fluctuations of the radiation field, they must be still observable today in the microwave cosmic background radiation. They could be indeed detected, although they are very small (see Sect. 6.1.2 and Figs. 6.6 and 6.7). We can therefore observe today the local matter density fluctuations at the time of the decoupling, i.e. about 400.000 years after the big bang.

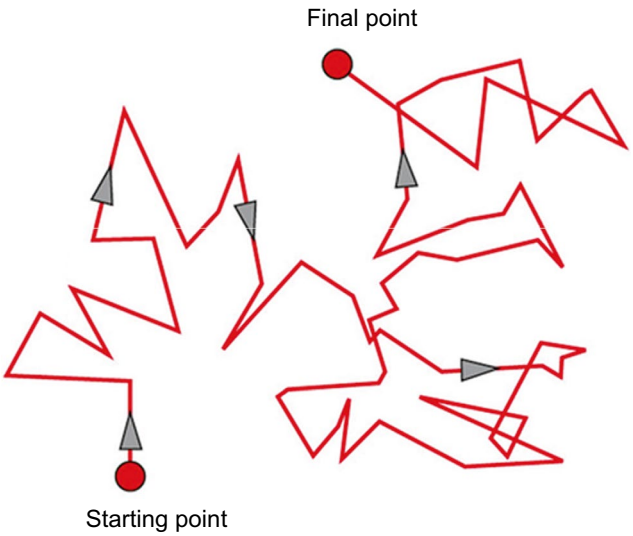


Fig. 6.21 Statistical motion of photons due to scattering by free electrons

As will be discussed later, such fluctuations can be due to microscopic quantum fluctuations in the very early stage of the universe before they have been blown up during the inflationary phase (see Sect. 6.3.12).

Detailed quantitative calculations have, however, shown, that the density of the visible matter is by far not sufficient to explain the observed structure formation. The matter must have a non-visible portion which is called "**dark matter**". It does not interact with electromagnetic radiation, and therefore, it is not visible. It does interact with the normal matter by gravitation. Its share must be essentially larger than the visible matter (see Sect. 6.6).

6.3.7 The Formation of Globular Clusters and Galaxies

If the mass density and temperature of neutral matter, including the dark matter, surpasses the **Jeans Criterion** (see Sect. 5.2.1), the collapse of the matter cloud can occur. If local density inhomogeneities are present, local condensation centers appear toward which the matter collapses. In these centers, stars are born and the many condensation centers form a large star cluster. This condensation takes place about $(2–3) \cdot 10^7$ years after the big bang, at a time when the mean matter density was about 10^6 times higher than today. It is assumed that these density fluctuations are caused by microscopic quantum fluctuations at times $t < 10^{-35}$ s before the inflation of the universe started. These microscopic fluctuations were then inflated to macroscopic size during the inflation period. They appear for the visible matter as well as for the dark matter. In Fig. 6.18, the different development stages of the universe are graphically summarized.

6.3.8 The Age of the Universe

Inserting into Eq. (6.31) the critical density ϱ_c from (6.32), the total energy $E = E_{kin} + E_{pot}$ of all galaxies with mass m, receding from us, becomes zero. We get from (6.29) and (6.30)

$$\frac{dr}{dt} = \left(\frac{2G \cdot m}{r}\right)^{1/2}. \tag{6.55}$$

Integration gives with $r(0) = 0$:

$$\left(\frac{2}{3}\right) r^{3/2} = (2Gm)^{1/2} \cdot t. \tag{6.56}$$

Today ($t = t_0$), we have the relation

$$\left(\frac{dr}{dt}\right)_{t_0} = v_0 = H_0 \cdot r = \sqrt{\frac{2Gm}{r}}. \tag{6.57}$$

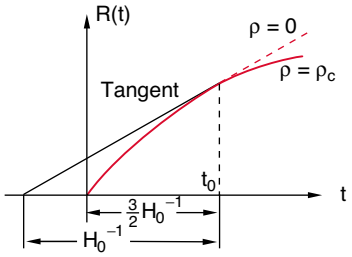


Fig. 6.22 Definition of Hubble time

$$\Rightarrow T_0 ==\Rightarrow 2/3H_0. \tag{6.58}$$

Inserting the numerical value of the Hubble constant $H_0 = 70\text{km/s} \cdot \text{Megaparsec}$, we get the age of the universe

$$t_0(\varrho_c) = 9.2 \cdot 10^9 \text{years}.$$

For $\varrho < \varrho_c$, the recession velocity of the galaxies becomes larger because the attractive gravitational force is smaller. For $\varrho = 0$ the velocity would be constant because no retarding force is present. This gives for the age of the universe

$$t_0(\varrho = 0) = 1/H_0 = 13.8 \cdot 10^9 \text{years}.$$

The time $t_0(\varrho = 0) = 1/H_0$ is called the **Hubble time** (Fig. 6.22).

For $\varrho < \varrho_c$, the age of the universe must be, according to this simple "Newton model" within the limits

$$\left(\frac{2}{3}\right) H_0^{-1} < t_0 < H_0^{-1} \Rightarrow 9.2 \cdot 10^9 \text{a} < t_0 < 13.8 \cdot 10^9 \text{y} \tag{6.59}$$

For $\varrho > \varrho_c$, the age t_0 becomes $t_0 < 2/3H_0^{-1}$.

Note:

Since the experimental value of H_0 is only certain within the error limits $67\text{km/(sMpc)} < H_0 < 75\text{km/(sMpc)}$, the Hubble age of the empty universe ($\varrho = 0$) calculated from these experimental values, is between $(12.5 < t_0 < 14.4) \cdot 10^9$ years and for $\varrho = \varrho_c$ between $(8.3 < t_0 < 9.6) \cdot 10^9$ years.

Independent experimental results, (e.g. from the detailed measurements of the cosmic background radiation, or from the age of the oldest stars) give for the age of the universe the more realistic result $t_0 = 13.7 \cdot 10^9$ years.

One of the reasons for the discrepancy with the earlier results is the fact that, in the earlier estimations, the existence of the dark matter, which increases the gravitational force, had not been taken into account. Furthermore, the dark energy, which results in an accelerated expansion of the universe after a certain time t_d was not included in the estimations.

From independent measurements, it is known that the age of the oldest stars in globular clusters, which contain mostly hydrogen but nearly no heavier elements, is larger than $12 \cdot 10^9$ years. Since the universe must be older than these stars, their age gives a lower limit for t_0. Model calculations of the early universe (see Sects. 6.3.3–6.3.6) give a realistic value of the time t when the first stars could have been born as $t = 10^8$ -10^9 years after the big bang. This leads to the age of the universe $t_0 = 13.7$ years, mentioned above. This value had been regarded in former times as the most reliable value of the age of the universe. The Hubble age derived from the Hubble constant $H_0 = 70\text{km/(s} \cdot \text{megaparsec)}$ gives only for $\varrho = 0$ the age $t_0 = 1/H_0 = 13.6 \cdot 10^9$ years. However, the assumption $\varrho = 0$ is unrealistic. Therefore, the Newton model outlined above has to be improved. It turned out that this can be done by the Friedmann equations, where a "**cosmological constant Λ**" is introduced, as will be discussed in the next section.

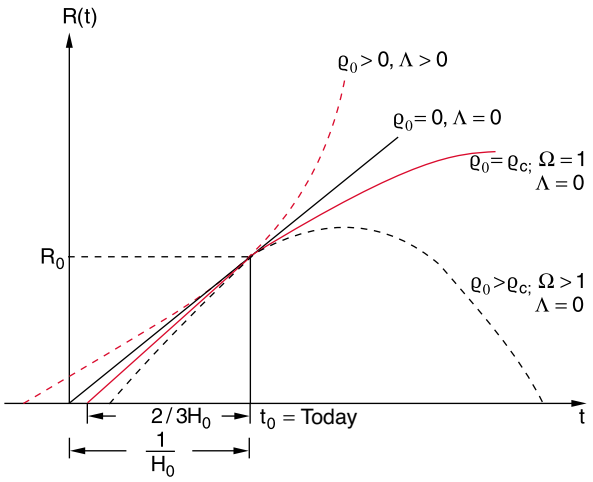


Fig. 6.23 Scale parameter R(t) for different mass densities ρ for a given value of the Hubble constant $H_0 = H(t_0)$

This constant Λ can be interpreted as a repulsive force, which acts against the attractive gravitational force and which causes a decrease in the deceleration and can even result in an accelerated expansion, dependent on the magnitude of Λ. The extrapolation of the values $R(t_0)$ and $(dR/dt)_{t_0}$ measured today toward $t \to 0$ gives a larger age of the universe (Fig. 6.23).

6.3.9 Friedmann Equations

We start with the equation of motion

$$m \cdot \frac{d^2 r}{dt^2} = -(4/3)\pi G \rho \text{mr} \tag{6.60}$$

for a galaxy with mass m at the distance r from the earth, which recedes with the velocity $v = H_0 \cdot r$ and is decelerated by the attractive gravitational force.

The introduction of the scale parameter R gives with $r = (R/R_0)r_0$

$$\frac{d^2R}{dt^2} + \frac{4}{3}\pi G\varrho(t) \cdot R = 0. \tag{6.61}$$

Since the total mass has to be preserved during the expansion, we get the condition

$$\varrho(t) \cdot R^3(t) = \varrho_0(t_0) \cdot R^3(t_0) \tag{6.62}$$

Eliminating $\varrho(t)$ in (6.61) and replacing it with the density $\varrho_0 = (R/R_0)^3 \cdot \rho(t)$ gives

$$\frac{d^2R}{dt^2} + \frac{C}{R^2} = 0 \text{mit} C = \frac{4\pi}{3} G\varrho_0 R_0^3. \tag{6.63}$$

Multiplication with dR/dt and integration from t_0 to t yields with

$$\vec{R} \cdot \dot{R} = \frac{1}{2}\frac{d}{dt}(R^2)$$

the **Friedmann equation**.

$$\left(\frac{dR}{dt}\right)^2 - \left(\frac{dR_0}{dt}\right)^2 - 2C\left(\frac{1}{R} - \frac{1}{R_0}\right) = 0, \tag{6.64a}$$

Which can be written with $\dot{R}_0 = H_0 \cdot R_0$ as

$$\left(\frac{dR}{dt}\right)^2 - \frac{2C}{R} + k = 0 \tag{6.64b}$$

with

$$k = \left(-H_0^2 + \frac{8\pi G}{3}\varrho_0\right)R_0^2 = (\varrho_0 - \varrho_c)\frac{8\pi G}{3}R_0^2, \tag{6.64c}$$

where the right side follows from (6.32) and (6.63).

The parameter k is a measure of the difference between the density ϱ_0 today and the critical density ϱ_c defined by Eq. (6.32). Dividing (6.64b) by R^2, one obtains the **Friedmann equation** in the form

$$(\dot{R}/R)^2 = (8\pi G/3)\varrho - k/R^2. \tag{6.64d}$$

In order to obtain realistic solutions of the Friedmann equation, one has to know, how the density $\varrho(t)$ changes with time during the expansion. This knowledge can be achieved with the first law of thermodynamics

$$dU + pdV = TdS \tag{6.65}$$

where we can express the "inner energy U" by the relation $E = mc^2$ as the mass energy of the universe

$$U = (4/3)\pi R^3 \varrho c^2 \tag{6.66}$$

Since R(t) as well as $\varrho(t)$ change with time, we have to use for the differentiation dU/dt) the product rule and get

$$\frac{dU}{dt} = \frac{dU}{dR} \cdot \frac{dR}{dt} = \left[4\pi R^2 \varrho c^2 + (4\pi/3)R^3 c^2 \frac{d\varrho}{dt}\right]\frac{dR}{dt}. \tag{6.67}$$

The volume change is

$$dV/dt = 4\pi R^2 dR/dt. \tag{6.68}$$

We assume, that the expansion proceeds reversible, i.e. the entropy S does not change with time, $\rightarrow dS/dt = 0$.

Then we get from (6.65)–(6.68) the equation of state

$$d\varrho/dt + 3(\dot{R}/R)(\varrho + p/c^2) = 0. \tag{6.69}$$

In a homogeneous universe, the pressure p does not cause any acceleration because it acts uniformly from all sides. If we set p = 0 in Eq. (6.69) the equation simplifies to

$$d\varrho/dt = -3(\dot{R}/R)\varrho \rightarrow \varrho(t) \sim R(t)^3. \tag{6.70}$$

The solutions of the Friedmann equation depend on the value of the parameter k, i.e. on the difference $a_0 - a_0 a_0 - a_0$

1. $k = 0 \rightarrow \rho_0 = \rho_c$.

The Eq. (6.64) is then elementary integrable and yields

$$R(t) = \left(\frac{3H_0}{2}\right)^{2/3} \cdot R_0 \cdot t^{2/3}. \tag{6.71}$$

The scale parameter R(t) and with it the radius r of the universe increases with $t^{2/3}$ and strives with decreasing slope $dR/dt \propto t^{-1/3}$ toward $R = \infty$ (Fig. 6.23).

2. Empty Universe $\rho_0 = 0 \rightarrow C = 0$; $k = -H_o^2 R_0^2$.

The solution of (6.64b) gives:

$$R(t) = H_0 R_0 t, \tag{6.72}$$

This means a linear increase of the scale parameter with time. This is obvious because there is no matter which decelerates the expansion due to gravitational attraction. Only the space expands.

3. K < 0, I.E. $\rho_0 < \rho_c$

Now the expansion is less decelerated, because of the smaller gravitational attraction. The slope dR/dt of the curve R(t) is always larger than for the case 1). The solution of (6.48) is now

$$R(t) = R_0 \cdot \cosh(ut - 1), \tag{6.73}$$

where u is a parameter depending on ϱ.

4. K > 0 I.E. $\rho_0 > \rho_c$

The gravitational force is now strong enough to stop the expansion. From (6.64b), we can conclude the maximum scale parameter R(t), which is reached for $dR/dt = 0$ and becomes

$$R_{\max} = \frac{2C}{k} = \frac{\varrho_0}{\varrho_0 - \varrho_c} \cdot R_0. \tag{6.74}$$

The solution of (6.64b) is

$$R(t) = R_0(1 - \cos ut). \tag{6.75}$$

Example

$\varrho_0 = 2\varrho_c \Rightarrow R_{\max} = 2R_0$. The expansion converts into a contraction at two times the present scale parameter R_0. ◀

Remark In the literature, the ratio of the present density ϱ_0 to the critical density ϱ_c is often denoted by the parameter Ω:

$$\Omega = \varrho_0/\varrho_c. \tag{6.76}$$

The case $\varrho_0 = \varrho_c \rightarrow \Omega = 1$ corresponds to k = 0. The empty universe $(\varrho_{\beta 0} = 0) \rightarrow \Omega = 0$ corresponds to $k = -H_0^2 R_0^2$. For $\varrho_0 < \rho_c \rightarrow \Omega < 1$ and for $\varrho_0 > \varrho_c \rightarrow \Omega > 1$.

The treatment in the framework of the General Relativity theory with the line element $ds^2 = c^2 dt^2 - dl^2$ and the metric (6.18) gives the Friedmann–Lemaitre Equation

$$\left(\frac{dR}{dt}\right)^2 - \frac{2C}{R} + k + \frac{\Lambda}{3} \cdot R^2 = 0$$

$$\Rightarrow \left(\frac{\dot{R}}{R}\right)^2 + \frac{k}{R^2} - \frac{8\pi G}{3}\varrho(t) = -\frac{\Lambda}{3}, \tag{6.77}$$

which contains the additional term of the cosmological constant Λ with the dimension $1/s^2$. For $\Lambda = 0$ the Eq. (6.77) reduces again to (6.64b).

A positive cosmological constant $\Lambda > 0$ corresponds to a repulsive force that acts against the attractive gravitational force. Within the framework of the quantum field theory, the cosmological constant can be regarded as a measure of the energy density w_{vac} of the vacuum. Introducing formally the mass density $\varrho_{vac} = w_{vac}/c^2$, the relation

$$\Lambda = 8\pi G \cdot \varrho_{vac} \tag{6.78}$$

Can be derived. Inserting this into (6.64), one gets instead of the density ϱ the lower density $(\varrho - \varrho_{vac})$.

The energy density of the vacuum acts as an additional repulsive force, which lowers the action of the attractive decelerating gravitational force. Calculations of the vacuum energy density give, however, a value, which is larger by many orders of magnitude than the value which is compatible with the observations. In recent years, theories have been developed which assume instead of the vacuum energy a "dark energy" (often called **quintessence**) which fills the complete space. The cause of this energy and its physical meaning are still completely unknown and long discussions try to clear this serious flaw. An important point in favor of a repulsive force is the experimental observation of the acceleration of the expansion, which is generally accepted. Since the vacuum energy density should be constant during the expansion, the gravitational energy decreases. There is a time when $\varrho - w_{vac}/c^2 < 0$ and the dark energy dominates over the gravitational energy. From this time on the expansion should be accelerated instead of decelerated. In Fig. 6.24, the expansion dR/dt is illustrated schematically with the inclusion of the dark energy. The curvatures of the curve R(t) are exaggerated in order to clarify the transition from a decelerated to an accelerated universe. The extrapolation of the slope $dR/dt = H_0 \cdot R$ measured today as $t = 0$ yields a larger value of the age of the universe.

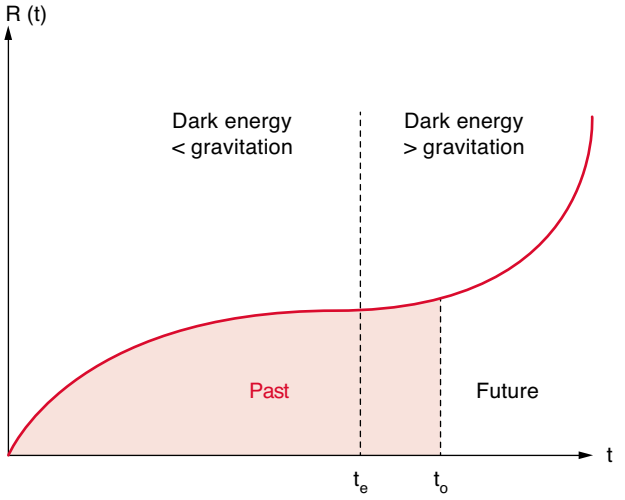


Fig. 6.24 Expansion of the universe with consideration of the dark energy

Up to now, the numerical value of the cosmological constant Λ cannot be measured with sufficient accuracy. As an upper limit, $|\Lambda| < 10^{-51} s^{-2}$ is accepted. Therefore, in most calculations, one sets $\Lambda = 0$.

6.3.9.1 The Redshift

The observations of Hubble have shown that the spectral lines, emitted from far distant galaxies, are shifted to longer wavelengths. This redshift

$$\frac{\Delta\lambda}{\lambda_0} = \frac{(\lambda - \lambda_0)}{\lambda_0} = \frac{\lambda}{\lambda_0} - 1 \tag{6.79}$$

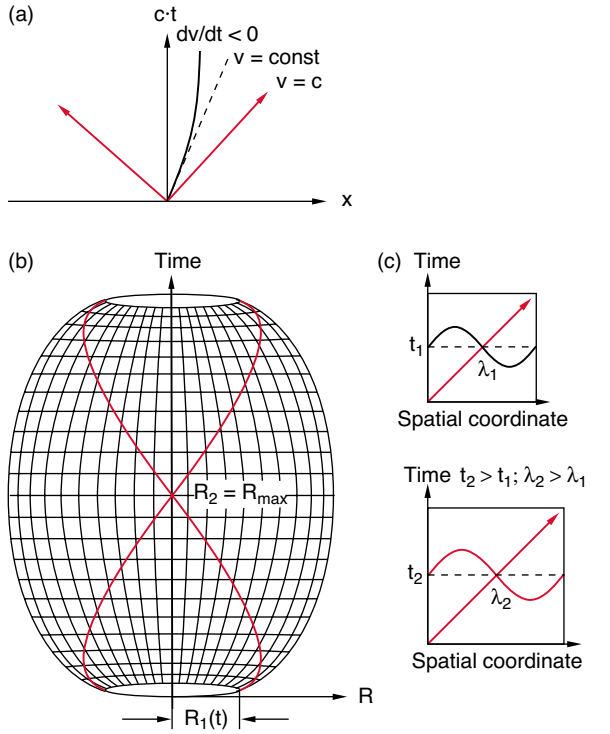


Fig. 6.25 **a** Minkowski diagram for a Euclidean metric with the world lines of photons (red curve) and galaxies (black curve). **b** Space–time metric of curved space. **c** Expansion of the space as the real cause of the redshift illustrated by small squares

$\Delta\lambda$ = observed wavelength λ_0 minus emitted wavelength λ had been explained as Doppler shift which means that the galaxies move with the radial velocity $v_r = c \cdot \Delta\lambda/\lambda_0$ away from us. The relative wavelength shift, called the **redshift**, then becomes for $v \ll c$ and can be defined by the number

$$z = \frac{\Delta\lambda}{\lambda_0} \approx \frac{v_r}{c} = H_0 \cdot r/c \qquad (6.80)$$

assuming the Hubble relation $v = H_0 \cdot r$.

Even if

- A more detailed consideration shows, however, that the real cause of the redshift is not the escape velocity of the galaxies but the expansion of the space, i.e. the increase of the scale parameter R(t).

This can be seen as follows:

In a space–time diagram, which describes a flat universe (Fig. 6.25a), the world lines of a photon are the red straight lines that form the angle 45° against the axes (see Vol. 1., Sect. 6.3.6).

The world line of a galaxy, which moves with the constant velocity v<c, is the black dashed straight line. If the velocity v of a galaxy decreases with time, its world line is the curved black line.

In a curved space, where the radius of curvature R(t) changes with time, the situation is different (Fig. 6.25b). This is illustrated by the example of a closed universe ($\varrho > \varrho_c$). The time axis is the vertical center line. The world lines of the galaxies are the red curved lines.

Drawing a three-dimensional grid it would show in a flat universe a net of squares, which are distorted in a curved space. In Fig. 6.25b of the time–space diagram of a closed universe, the radius of curvature R(t) increases until it reaches the maximum value $R_{max}(t_2) = \varrho_0 R_0/(\varrho_0 - \varrho_c)$ and then decreases again, i.e. the universe contracts again. The world line of a photon proceeds from one corner of the distorted squares to the opposite corner but appears in the total grid as a curved line because the whole space is curved. A galaxy, which is located in one corner of the square, participates in the expansion of the whole space without moving into the direction of the space coordinate. Its distance from us had been r_1 at the time t_1 and has increased to $r_2 > r_1$ at the time t_2 because the sidelength D of the square has increased during the expansion of the universe proportional to the scale parameter R(t).

We can therefore say that the space is expanding and this causes an increase in the distance to the galaxies. Remember the example of the small disks glued on the surface of the inflatable balloon, they cannot move on the surface but their distance increases when the balloon is inflated.

During this inflation of the space, all lengths, and therefore, also the wavelengths of light increase. This is illustrated in Fig. 6.25c. Assume that a galaxy has emitted at the time t_1 light with the wavelength λ_1, we choose our grid so small that the sidelength of a square just equals λ_1. The light travels on a diagonal path through the square. The time of flight is $\Delta t = \lambda_1/c$. At a later time $t_2 > t_1$, the grid widens because R(t) and D(t) have increased. Since the expansion time in the mass-dominated universe increases as $t \propto R^{3/2}$ (6.42), the sidelength D(t) of our square also increases with time $D(t) \propto t^{2/3}$. Since $D(t) \propto R(t)$, we get the relation

$$\frac{R_1(t_1)}{R_2(t_2)} = \frac{\lambda_1}{\lambda_2} = \left(\frac{t_1}{t_2}\right)^{2/3}. \qquad (6.81)$$

This means that light, which was emitted with the wavelength λ_1 at the time t_1 from the galaxy, has the wavelength λ_0 when it is received on earth at the time t_0.

With the relative wavelength shift

$$z = \frac{\Delta\lambda}{\lambda_0}$$

we can write (6.81)

$$\frac{\lambda_0}{\lambda} = (1+z) = \left(\frac{t_0}{t_e}\right)^{2/3}. \qquad (6.82)$$

Therefore, the momentary distance r (t_0, z) of the galaxy emitting at $t=t_1$ radiation with the wavelength λ_1, which has suffered the wavelength shift z, had the distance r_1 at the time t_1 of the emission but is now away from us at the distance

$$r(t_2, z) = \frac{2c}{H_0}\left[1 - \frac{1}{\sqrt{1+z}}\right] + r_e. \qquad (6.83)$$

The pathlength which the light has traveled is, however, only

$$l(z) = c(t_0 - t_e) = ct_0[1 - (t_e/t_0)]$$

$$= \frac{2c}{3H_0}\left[1 - \frac{1}{(1+z)^{3/2}}\right]. \qquad (6.84)$$

▶ This shows that for larger values of z a galaxy is at the time t_0 of the detection much farther away from us than the pathlength traveled by the light between emission and detection. The larger the redshift is, the earlier the time of the emission.

Galaxies with a large redshift allow us a glimpse of the state of the universe at a very early time

$$t_1 = t_0 - \frac{2}{3H_0}\left[1 - \frac{1}{(1+z)^{3/2}}\right]. \qquad (6.85)$$

Note: One must distinguish between the present distance of the galaxy at the time t_0 of the radiation detection and its distance at the time t_1 when the radiation was emitted. This is illustrated in Fig. 6.26.

The expansion velocity of the universe can be larger than the speed of light.

The conclusion that the measured redshift gives the velocity of a receding galaxy has to take care of some other reasons for the redshift:

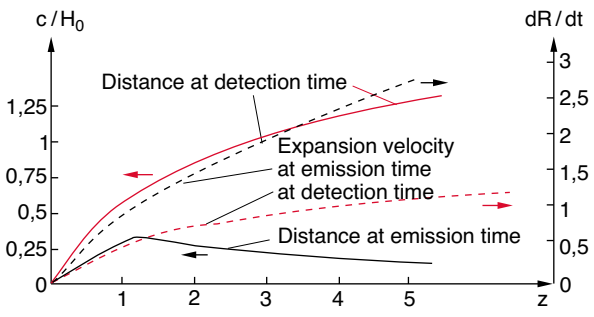


Fig. 6.26 Distance c/H_0 of a galaxy (solid curve) and escape velocity (dashed curve) for $\rho < \rho_c$ at the emission and the detection time of a signal as a function of the redshift z

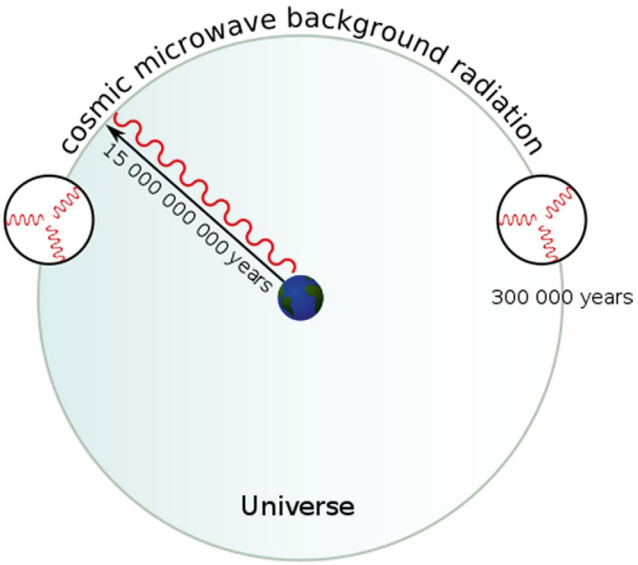


Fig. 6.27 Illustration of the horizon problem (https://en.wikipedia.org/wiki/Horizon_problem)

(1) The expansion of space discussed above
(2) The real Doppler shift observed when a star in our galaxy moves away from us with the velocity v is

$$z = \sqrt{\frac{(c+v)}{(c-v)}} - 1. \qquad (6.86)$$

This wavelength shift is, for example, used for the measurement of velocities of stars in our galaxy against the earth. Within the galaxy, the expansion of space can be neglected because the galaxy is held together by attractive gravitational forces, and therefore, does not noticeably expand. For far away galaxies with distances above 100 Mpc, the movements of a galaxy in a galaxy cluster can be neglected against the expansion of space. One has to take into account the different causes for redshifts:

1. The redshift of light in the gravitational field (see Sect. 6.6.2). If, for instance, light is emitted from the surface of a massive star with mass M, it suffers a redshift $\Delta\lambda/\lambda = G{\cdot}M/(Rc^2)$.
2. If the emitting object is far away, the light is scattered and partly absorbed when it passes through gas- and molecular clouds. The scattering cross section is $\sigma_{sctt} \propto 1/\lambda^4$. Blue light is more strongly scattered than the red light. Therefore, the line center of broad spectral lines is shifted toward the red.

It is, therefore, necessary to perform a careful analysis of the measurements of redshifts in order to determine the different contributions separately and to extract the wanted contribution with sufficient accuracy. Since all

measurements are made from the rotating earth moving in space, its motion has, of course, also to be taken into account.

Example

For the galaxies with the largest distances measured so far, the redshift is $z = 5, \Rightarrow t_1 = t_0/6^{3/2} = t_0/14.7$. With $t_0 = \left(\frac{2}{3}\right)/H_0$ and $H_0 = 73\text{km}/(\text{s} \cdot \text{Mpc})$ we get.

$$t_0 = 9.2 \cdot 10^9\text{y} \Rightarrow t_1 = 6.3 \cdot 10^8\text{y}.$$

The light was emitted at the early time of the universe. It is remarkable that less than 1 billion years after the big bang, galaxies with a large luminosity did exist [15]. The present distance r(z, t_0) of the galaxy according to (6.83) is

$$\text{R}(z, t_0) = 4.8 \cdot 10^9\text{pc} = 15.5 \cdot 10^9\text{ly},$$

While the light travels only the pathlength

$$L(z) = c \cdot (t_0 - t_1) = 8.68 \cdot 10^9\text{ly}\ .$$

Remark For such large values of z, the escape velocity v reaches nearly the velocity c of light. Therefore, instead of (6.82), one has to use the relativistic Doppler shift

$$z + 1 = \left(\frac{(1 + v/c)}{(1 - v/c)}\right)^{1/2}.$$

This changes the numerical values in the example.

6.3.9.2 So Far Unsolved Questions

There are some problems in Cosmology, which could not be solved satisfactorily up to now. These are:

(a) The horizon problem raises the question of why different parts of the universe that have no contact with each other, because they are too far apart, have nevertheless equal properties such as temperature, energy density and luminosity.
(b) The question of why our universe is flat according to our present knowledge. This is equivalent to the question, why among many possible initial conditions at the beginning of the universe just that one was realized where the density is equal to the critical density.
(c) How was it possible that from the initial state of the universe with homogeneous and isotropic energy, density structures could be developed, such as galaxies and stars?

These three questions are answered by the model of the inflationary universe proposed by *Alan Guth*. However, even when accepting this model, there remain some open questions and there are discussions on how to improve this model or to invent alternative theories.

(a) The Horizon Problem.

We define the horizon at time t for an event that took place at the earlier time t_0 as the spherical shell with radius $r = c(t - t_0)$ and the center at the location of the event. At the time $t = 380.000$ years after the big bang, when the decoupling of matter and radiation took place, this radius was $r = 380.000$ light years. The 2.7K cosmic background radiation gives us a picture of the universe at that time, because from that time on the radiation could expand unperturbed.

As was shown in Sect. 6.3.9, the scale factor R(t) increases with time in the mass-dominated universe as $R(t) \propto t^{2/3}$. Assuming the age of the universe as $t_0 = 13.7 \cdot 10^9$ years, the scale factor at the time $t_1 = 380.000$ years had been

$$R(t_1) = \left(3.8 \cdot 10^5/13.7 \cdot 10^9\right)^{2/3} \cdot 13.7 \cdot 10^9\text{ly} \approx 8 \cdot 10^6\text{ly},$$

which is essentially larger than the horizon with radius $3.8 \cdot 10^5$ ly. Our present observable universe, which lies within the present horizon, had been at the time $t_1 = 380.000$ years much larger than the horizon at that time. Therefore, there had been large parts of the universe, which could not receive signals from points inside the horizon.

Example

Assume that we observe two galaxies at a distance of 10 billion light years from the earth but in opposite directions. The distance between the two galaxies is then 20 billion ly. They could not communicate with each other because their distance was too large and even at the time when they had been created, they could not exchange signals, because one of the galaxies was outside the horizon of the other galaxy. It is therefore strange that they have the same physical characteristics. ◀

(b) Is our Universe flat?

The question of why our universe is flat although this represents only one possibility among a huge number of other possibilities can be illustrated as follows:

For a flat universe, the density ρ equals the critical density ρ_c and the parameter $\Omega = \frac{\varrho}{\varrho_c} = 1$. The parameter Ω is related to the Hubble parameter H(t) and the scale factor R(t) through the equation

$$|\Omega - 1| = |k|/\left(R^2H^2\right), \tag{6.87}$$

where $k = (\varrho_0 - \varrho_c)(8\pi G/3)R_0^2$.

If $\Omega = 1$, the density ϱ remains for all times equal to the critical density ϱ_c, which means the universe is flat for all times.

For $\Omega > 0$ is $\varrho > \varrho_c$ and the normalized difference $(\varrho - \varrho_c)/\varrho_c$ increases with time. In the radiation dominated universe is $\propto t$, in the mass dominated universe it is $\propto t^{2/3}$ because $(R \cdot H)^2 \propto t^{-1}$ resp. $\propto t^{-2/3}$.

If presently $\varrho = \varrho_c$, it must have been also in the beginning of the universe $\varrho = \varrho_c$, which means there must have been a "fine-tuning" that among many possibilities for the density ϱ, just the one $\varrho = \varrho_c$ must have been exactly realized, because any tiny positive deviation from $\varrho = \varrho_c$ would have resulted in a steady increase of the difference.

Example

If we would measure today a difference $\varrho - \varrho_c = 0.1\varrho_c$ the difference $|\Omega - 1|$ were at the time of the decoupling of matter and radiation ($t_1 = 380.000$ years) $|\Omega - 1| < 10^{-5}$ and at the time of the nucleo-synthesis ($t = 1s$) even $|\Omega - 1| < 10^{-18}$! This demonstrates the phrase "fine-tuning" because a very tiny deviation from the condition $\varrho = \varrho_c$ would have a large effect at later times. It seems highly improbable that among many possible initial conditions, just this narrow range for the density ρ was realized. ◀

6.3.9.3 The Cosmic Inflation

There are several proposals for the solution of this problem. The one, favored by most cosmologists is the model of cosmic inflation proposed by Alan Guth [9].

His hypothesis assumes that the universe at the age of 10^{-35} s, when the temperature had fallen below the critical value of $T = 10^{28}$ K (above this value a quark-gluon plasma existed where all 4 interactions were unified, i.e. not distinguishable), expanded for a very short time (until 10^{-32} s) with an exponentially increasing expansion velocity at a constant energy density (Fig. 6.28). This exponential expansion can be derived from the Friedmann Eq. (6.64) as follows.

The transition from the quark-gluon plasma to separated hadrons and leptons is a phase transition, where energy is released. It is also called "symmetry-breaking" because hadrons and leptons can now be distinguished as different particles with different masses. Furthermore, the gravitation is separated from the other three interactions. The energy released at this phase transition provides the energy needed to keep the energy density of the universe constant in spite of the expanding space.

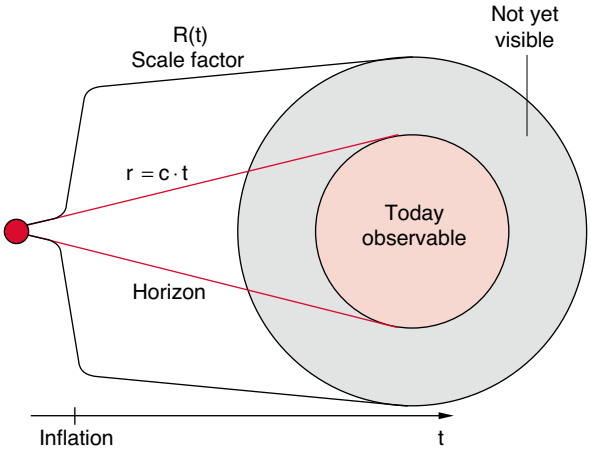


Fig. 6.28 Illustration of the inflationary expansion of the universe

Inserting $\varrho = \text{const.}$ into the Friedmann equation (6.64), one obtains with $C = \text{const.}$

$$dR/dt = a \cdot R \quad \text{with} \quad a \approx 10^{34}\text{s}^{-1}. \tag{6.88a}$$

Integration gives

$$R = R_0 \cdot e^{at} \tag{6.88b}$$

which represents an exponentially increasing scale parameter R.

A. Guth assumes that R doubles every 10^{-34}s. This gives an expansion velocity of the space which is by far larger than the velocity of light. However, this does not contradict the postulates of the relativity theory, because here it is not the velocity of matter, but only that of space and it is the space that expands not matter. The end of the inflation which occurs at $t = 10^{-32}$s is determined by the total consumption of the energy released during the phase transition. During the inflation phase, the scale parameter has increased from $R_1(10^{-35}\text{s}) = 10^{-35}$ light seconds $= 3 \cdot 10^{-27}$m to $R_2(10^{-32}\text{s}) \approx 1$ m (Fig. 6.29). This corresponds to an expansion velocity of $v = 1\text{m}/10^{-32}\text{s} = 10^{32}$ m/s which is larger than the velocity of light by the factor $3 \cdot 10^{23}$.

In the inflation model, the initial scale parameter at the beginning of the inflation is assumed to be smaller by a factor of 10^{40}–10^{50} than in the Standard Model. However, this does not imply that the whole universe had such a small volume. It is sufficient to assume that only in a small part of the university the energy density had a higher value due to quantum fluctuations. This part could then inflate due to its higher energy density. After the inflation, this part covers the whole universe because the scale parameters have now the same value as in the standard model. After 10^{-32}s, the inflation passes over into the normal expansion

according to the Standard Model, where the matter density is $\varrho \propto 1/R^3$ and the energy density $w \propto 1/R^4$ because now the released energy of the phase transition has been completely expended.

How can this inflation model solve the horizon problem?

Thermal equilibrium can only occur within the event horizon because it demands the exchange of energy and momentum, which can only be realized with velocities $v \leq c$. The radius of the horizon at the time 10^{-34}s the value $r = 10^{-34}$ light seconds $\approx 3.10 \cdot 10^{-26}$m. The scale parameter had, however, according to the inflation model only the value $R = 10^{-40}$m, i.e. the total space lies inside the horizon. Therefore, the whole volume of the universe is at thermal equilibrium. At the end of the inflation, the horizon has increased from 10^{-34} light seconds to 10^{-32} light seconds, but the scale parameter has increased to more than 1 m. Now the horizon is much smaller than the scale parameter. Nevertheless, the space outside the horizon keeps the remembrance of the thermal equilibrium at the time before the inflation, because the expansion was so fast, that the thermal equilibrium was not lost.

In the following time, the horizon increases with the velocity of light, and therefore, is much faster than the scale parameter. At the present time, it has reached the scale parameter of the observable universe.

The assumption that we live in a flat universe ($\varrho = \varrho_c$) is explained by the inflation model with the enormous expansion of space during the inflation phase where any possible initial curvature of space would have been flattened.

A modification of the inflation model assumes that during the inflation, the scale parameter had increased far beyond that of the standard model and is today larger than the radius of the horizon. This means that there are parts of the universe, which are inaccessible for us because we cannot see them (Fig. 6.29b). We see only a small part of the whole universe, which appears flat to us similar to the fact that we regard the water surface of a lake as flat, although the earth's surface is, of course, curved.

This modified model answers the three questions above if we agree with the assumptions of this model:

(a) The universe observed today derives from the volume at $t = 5 \cdot 10^5$ y which was enclosed by the horizon, where thermal equilibrium was established. Therefore, the present homogeneous and isotropic mass distribution is the consequence of the former thermal equilibrium.
(b) The universe observed today represents only a small section of a much larger curved space. This section appears flat because its dimensions are very small compared to the radius of curvature.
(c) The formation of structures in the universe, such as galaxies and stars, is caused by quantum k fluctuations in the very early stage of the universe, before the inflation started. These microscopic fluctuations were greatly expanded to macroscopic size by inflation.

Note: that the model of the inflation universe is still just a hypothesis that has not yet been sufficiently proved by experiments and observations, although it can answer the fundamental questions presented above. It is not contradictory to the standard model but represents a supplement. it is not clear whether the energy released during the phase transition is sufficient to drive the inflation and fill the enlarged volume with constant energy density, or whether the rare additional energy sources could drive the inflation.

The inflating universe which expands with a velocity, surpassing the velocity of light by many orders of magnitude, does not contradict the special relativity, because it is not the matter that inflates but only the space.

For more detailed representations of the standard model and all phenomena related to it, the reader is referred to the extensive literature [12–22].

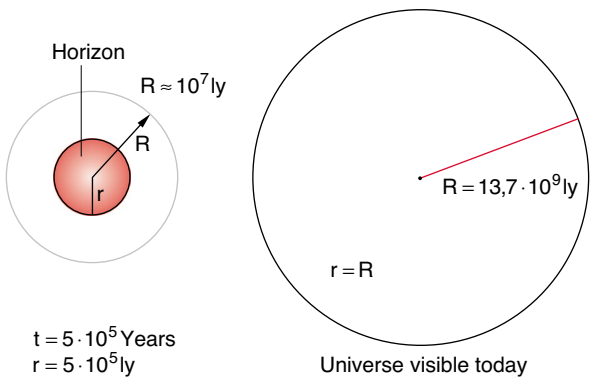


Fig. 6.29 **a** The scale factor R(t) before, during and after the inflationary phase. **b** The scale factor R(t) at the time $t = 5 \cdot 10^5$ years and today

6.4 Formation and Structure of Galaxies

Only recently a more detailed perception about the formation of early structures in the universe could be developed, supported by measurements of tiny deviations from the isotropy of the cosmic background radiation. The questions that have bothered astronomers for many years were: How can structures be formed already at relatively early times in a nearly homogeneous and isotropic universe and how are such structures develop into galaxies and stars? Only the assumption of the existence of dark matter could give satisfactory answers.

6.4.1 Dark Matter and Formation of Spatial Structure of the Matter Density

Dark matter is an until now unknown form of matter that interacts with each other and with visible matter only by gravitation and may be by weak interaction, but not by electromagnetic interaction. It, therefore, does not absorb or emit photons and is not directly visible but can be inferred by its gravitational interactions with the visible matter. According to our present knowledge, it is essentially responsible for the formation of mass condensation in the early universe after the decoupling of matter and radiation.

While the baryonic and leptonic matter before the decoupling of matter and radiation was hindered by the radiation pressure to clump together, the dark matter did not have these restrictions because it did not interact with the radiation. Therefore, it could agglomerate at locations with microscopic matter condensations, thus forming local density structures. After the decoupling of matter and radiation, the radiation pressure disappeared and also the visible matter could be attracted by these local density maxima and enforced them.

The question is now, where these microscopic density structures come from?

According to our present knowledge, there exists in the very early universe ($t < 10^{-36}$ s) local quantum fluctuations due to the uncertainty relation $\Delta r \cdot \Delta p \geq \hbar$, or $\Delta E \cdot \Delta t \geq \hbar$ which cause random energy fluctuations related to local density fluctuations ($\Delta m = \Delta E/c^2$) within volumes with dimensions of $\left(10^{-8}\text{m}\right)^3$. These microscopic density fluctuations are blown up during the inflation phase to macroscopic fluctuations. For the dark matter they are not suppressed by the radiation pressure, because it does not interact with the radiation. They can, therefore, develop even before the decoupling of matter and radiation. The density maxima attract through the gravitational interaction of the baryonic matter after the decoupling time and enhance the regions of higher matter density. These density maxima become the nuclei for the formation of galaxies.

The density fluctuations can be quantitatively treated by introducing the density contrast

$$\Delta_r(r) = (\varrho(r) - \overline{\rho})/\overline{\rho}, \tag{6.89}$$

where r is the location where the density $\varrho(r)$ is measured and $\overline{\rho}$ is the mean density in the whole universe. The lower index r indicates the magnitude r of the volume where the density is measured. Measuring the density contrast for volumes of different sizes, the mean fluctuation σ can be defined for n measurements as

$$\sigma = (1/n)\sqrt{\Sigma}\delta_r(r)^2. \tag{6.90}$$

In Fig. 6.30, the mean fluctuation σa is plotted as a function of the volume size.

The important point is now the following:

These density fluctuations of matter shortly before the decoupling time also reflect the fluctuations of the energy density of the cosmic background radiation, because radiation and matter are coupled. After the decoupling, the radiation density is no longer influenced by the matter density fluctuations. Therefore, measurements of the small anisotropy of the present background radiation allow the conclusion of matter density fluctuations at the time $t_1 = 380.000$ years of the decoupling.

Measuring the anisotropy of the background radiation over large angular ranges, one gets the density fluctuations within large volumes. With a better angular resolution also the density fluctuations in smaller volumes can be determined.

Detailed measurements of the anisotropy of the cosmic background radiation at different angular resolution gives us information about the density fluctuations of matter in the universe at the time of decoupling of matter and radiation. It turns out that the universe is all the more homogeneous and isotropic the larger the regarded volume is for which the mean density is measured.

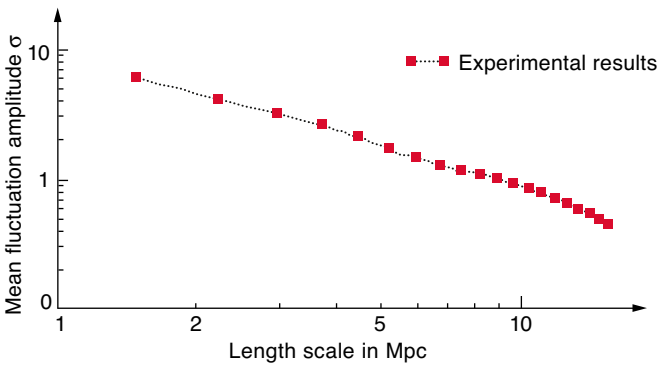


Fig. 6.30 Mean fluctuation strength as a function of the regarded volume

6.4.2 What Are Galaxies?

By the end of the eighteenth century, astronomers noticed that besides stars which appeared as point-like light sources, there were also diffuse luminous objects seen in the sky, which were called "nebulas". The most established one was the *Andromeda nebula* (Fig. 6.31), which can be seen with the naked eye. After the development of large telescopes, it was recognized that many of these nebulas were really an accumulation of millions to billions of stars. They are called today *galaxies* instead of nebulas, in order to distinguish them from true nebulas, which represent extended regions of luminous gas or dark areas of dust and absorbing molecules..

The name galaxy is the Greek name for "milky way" (Sect. 5.1). By observing the sky on a clear summer night from a dark observation point unperturbed by background radiation, one can see an elongated shimmery milky-white band *called the milky way* (Fig. 5.1). We will see in the next section that this milky way is caused by a large star density in our galaxy, because one looks when observing the milky way into the direction of the central plane of our flat galaxy where the star density is maximum. It represents the symmetry plane of our galaxy. Therefore, the name *milky way* has been extended to denote all stars of our galaxy.

The galaxies known so far are listed in catalogues. The older catalogue is the Messier catalogue named after the astronomer Charles Messier 1730–1817), and the newer one is the New General Catalogue of Nebulas and Clusters of Stars edited in 1888 by J. L. E. Dreyer (1852–1926). The galaxies are denoted by M+number or by NGC+number. For example, the Andromeda Galaxy is named M31 or NGC 224.

More detailed descriptions of the formation and structure of galaxies can be found in [61–67].

6.4.3 Different Types of Galaxies

Our universe contains according to a realistic estimation more than 10^{11} galaxies. Only about 10% of them have been observed, while the other 90% are too far away or too faint to be observable. From this huge number, only about 500.000 have been investigated in more detail. Each of these galaxies contains about 10^9–10^{12} stars. This implies that our universe encloses the inconceivable large number of more than 10^{22} stars.

The distances between the galaxies are generally large compared to their diameter. For example, our neighbor galaxy, the *Andromeda-nebula* M31 (Fig. 6.30) is 650kpc, which is two million lightyears away from us, while its diameter is only 60 kpc. One may therefore regard the galaxies as islands in the huge cosmic space. Often congregations of many galaxies are found, which are called **galaxy clusters** and which contain up to 2500 galaxies.

One example of a small galaxy cluster is the *local group*, which contains besides our own galaxy the Andromeda nebula and about 30 additional small galaxies. An example of a large galaxy cluster is the *Virgo cluster* at a distance of 70 Million ly which contains about 2500 galaxies, or the Perseus cluster (Fig. 6.32) at a distance of 250 Million Ly,

Fig. 6.31 The Andromeda Galaxy, about 2.2 million lightyears away It can be seen with the naked eye. It has a similar spiral structur as our own galaxy

Fig. 6.32 The Perseus Cluster A426 includes about 1000 galaxies and belongs to the galaxy clusters with the largest number of galaxies. In its center, mainly elliptical and SO galaxies are found. (J. Brunzendorf Sternwarte Tautenburg)

which contains about 500–1000 Galaxies. The total mass of the Perseus cluster is about $2 \cdot 10^{15} M_{\odot}$.

The distances of the galaxies had been for a long time a matter of intense disputes. They can be estimated with the help of pulsating variable stars (see Sect. 5.8.1), because there is a relation between the pulsation period and absolute luminosity which allows the determination of distances by comparing the measured apparent luminosity with the calculated absolute luminosity, after having calibrated the method by known luminosities of cepheids in our own galaxy. Nowadays, the distances have been measured with great accuracy with the astrometric satellites HYPARCOS and GAYA (see Sect. 6.5x).

In order to learn more about the formation mechanism and the age of galaxies, it is useful to study their structure in more detail. The comparison of the observed structures of other galaxies can give us some hints about the structure of our own galaxy.

Already 1936 E. Hubble had recommended a classification of galaxies according to their structure, which is in principle today still valid. It distinguishes between elliptical, lens-like, Spiral- and irregular galaxies. The spiral galaxies are further divided into normal and barred spiral galaxies.

Initially, this Hubble sequence had been regarded as the time development of galaxies, where the elliptical galaxies represented the starting form from which later the spiral galaxies developed. Based on the observation of many thousand galaxies, it became clear that these different structures do not reflect the time sequence in the development of galaxies. It was also found that many intermedium structures between the ones classified by Hubble. This demanded a finer classification (Fig. 6.33). For example, the elliptical galaxies are now divided into 8 subclasses E_0– E_7, where the index $n = 10(1 - b/a)$ for the subclass E_n is a measure for the ellipticity depending on the ratio b/a of semi-minor axis b to semi-major axis a.

Example

The image of an E_0 galaxy is a circle, and that of an E_7 is, with $b/a = 0.3$, an elongated ellipse. ◀

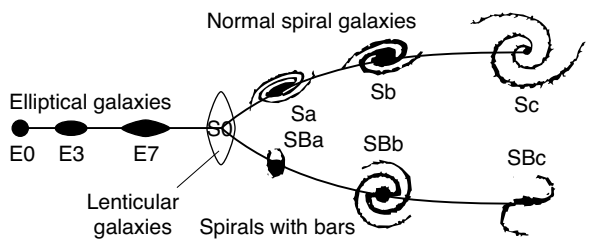


Fig. 6.33 Classification of the different galaxy types, starting from the Hubble sequence

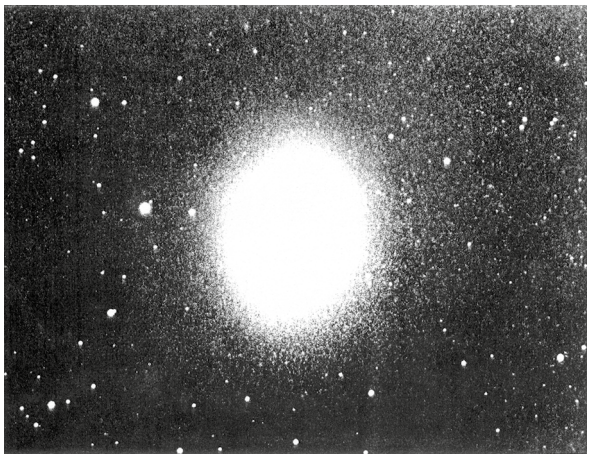


Fig. 6.34 The elliptical galaxy M32 of type E2. It is a small companion of the Andromeda galaxy

Note: Also, an elongated ellipse can appear circular if its major axis is inclined against the observation direction, and therefore, seems to be shorter. The projection onto a plane perpendicular to the viewing directions can appear as a circle if the inclination angle has the suitable value.

From many observations, it became clear that the elliptical galaxies are indeed elongated ellipsoids and not spheres.

Most of the large galaxies are elliptical galaxies. One example is M32, shown in Fig. 6.34 and in the upper right corner of Fig. 6.31. Their size and luminosity differ considerably and they contain between 10^8 stars (small galaxies) to 10^{13} stars (giant galaxies). Elliptical galaxies contain old stars, which originated about 10^{10} years ago. They comprise only a small amount of interstellar gas and dust. This implies that no new stars can be created. The main part of the luminosity is emitted by red giant stars, while the largest fraction of the total mass is provided by main sequence stars. The elliptical galaxies therefore appear reddish in color. The surface brightness decreases from the center to the edges. Its measurement gives information about the three-dimensional structure of the galaxy.

The rotational angular velocity of elliptical galaxies is mall, much smaller than the random eigen-velocities of the stars in the galaxy.

The spiral galaxies (Figs. 6.35 and 6.36) which represent about 75% of all observed galaxies have a more sophisticated structure than the elliptical galaxies. They show a central bulge with a structure and star composition similar to that of the elliptical galaxies. Around this galactic center, a flat disk with a spiral structure appears, which contains the majority of bright blue stars. The spiral arms there shine in bluish light. The diameter of the disk is about 30.000 pc but its thickness is only 700 pc. Many spiral galaxies including our own galaxy show a bright core region with the shape of a bar. They are called barred spiral galaxies.

Fig. 6.35 Pinwheel galaxy M101 or NGC5457 as an example of a normal spiral galaxy of type Sc (Photographed by the Hubble Space Telescope)

Also, the spiral galaxies are divided into subclasses SAa SAb SAc SAd (Fig. 6.33). The letter A means: no bar, i.e. normal spiral galaxies. The small letters give the opening angle of the spiral arms: (a for 10°, up to d for 25°). Galaxies of type SAc and SAd have largely open spiral arms (Fig. 6.35).

Barred spiral galaxies are denoted by a capital letter B (Fig. 6.36), this is because the formation of the bar takes a long time. Bar spiral galaxies are therefore older than the normal spiral galaxies. The bars are the location of star birth, that's why they shine bright.

Our own galaxy is a barred spiral galaxy of the type SBbc.

A very interesting information is achieved from measurements of the rotational curves, which means the dependence of the angular velocities ω(r) as a function of the distance r from the center of the galaxy. The measured opposite Doppler shift on both sides of the center allows the determination of the rotation velocities $v(r) = r \cdot w(r)$. Such direct measurements are only possible for galaxies with sufficiently large angular diameters, i.e. galaxies that are close to us. For those galaxies where only the total light from the whole galaxy can be measured without resolving the different radial parts, the Doppler-broadening of the spectral lines is measured. The analysis of the spectral line profile enables the conclusion of the angular velocities ω(r).

The escape velocity of the galaxy causes a Doppler shift of the line center.

The galaxies generally do not rotate as a rigid body but show a differential rotation, where the angular velocity depends on the distance r from the center. In Fig. 6.38, the rotation curves $v(r) = w(r) \cdot r$ are drawn for some spiral galaxies. It turns out that these rotation curves can be only satisfactorily explained if it is assumed that the mass of the galaxy is much larger than the visible mass and is not uniformly distributed of the galaxy, i.e. there must be dark matter which affects the rotation.

Our own galaxy, the milky way, is also a spiral galaxy of type SAbc, which represents a transition from a normal to a barred galaxy. In the spiral arms, large gas and dust clouds

Fig. 6.36 The barred spiral galaxy M 83 (NGC 5236) of type SAB (Photo: David Malin Anglo-Australian telescope)

Fig. 6.37 Barred spiral galaxy NGC 1300 of type SBbc in the constellation Eridanus, 67 million ly away (Hubble Space Telescope, NASA)

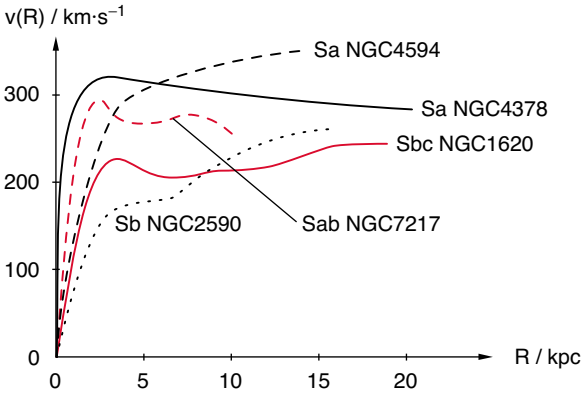


Fig. 6.38 Rotation Curves of some spiral galaxies (after V. C. Rubin, W. K. Ford, N. Thonnard: Astrophys. J.225. L107 (1978))

are present which are the regions where new stars are born (see Sect. 6.5y).

The third class of galaxies are **lenticular galaxies** classified as SO-galaxies. They lie in the Hubble classification scheme (Fig. 6.33) at the branching point between the elliptical and the two branches of spiral galaxies. They consist of a bright bulge, which resembles elliptical galaxies, and a surrounding disk-like structure that does not show spiral arms (Fig. 6.39).

The fourth class is irregular galaxies which have no specific form (Fig. 6.40).

About the formation of galaxies, only qualitative perceptions exist which are not safely confirmed. According to our present knowledge the dark matter played an essential role, because its density fluctuations did already exist before the decoupling of matter and radiation because the dark matter does not interact with the radiation and was therefore not hindered by the radiation pressure to clump together. The density maxima attracted after the decoupling time ($\approx$ 380.000 y after the big bang) the visible matter and formed gas and dust clouds with increased density which exceeded

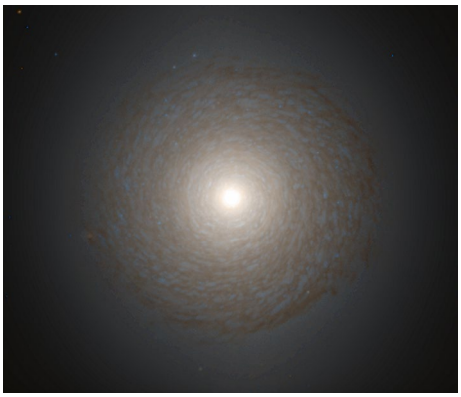


Fig. 6.39 Lenticular galaxy NGC 1387 in the Fornax Cluster (by Fabian RRRR—Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=63095419

the critical density of the Jeans Criterion. At first, small proto-galaxies were formed. By collisions between galaxies, larger galaxies developed (Fig. 6.41).

The birth of stars from local density maxima consumed the gas in the clouds, and therefore, today, elliptical galaxies no longer contain much gas. The present mass of the residual interstellar gas clouds is less than 10% of the total mass of a galaxy. The situation is different for spiral galaxies because they rotate. In the plane of the rotating disk, the gravitational attraction is compensated by the centrifugal force, while in the direction of the rotation axis perpendicular to the disk, matter can still fall onto the flat galaxy, thus providing sufficient gas densities to surpass the Jeans criterion so that new stars can be born. We will discuss this formation process of stars later by the example of our own galaxy.

In Table 6.1, the relative number of the different types among the brightest galaxies are listed.

6.4.4 Active Galaxies

In the foregoing sections, we have regarded "normal galaxies". During the last few years, many galaxies could be observed which do not fit into the general scheme of galaxies because of their luminosity and the spectral distribution of their emission. To this new class of galaxies belong the "**Seifert-galaxies**" named after their discoverer *Carl Seyfert,* and the **quasars** (quasi-stellar radio sources). The Seyfert galaxies which account for about 10% of all galaxies have a very bright central part (about 10^8–10^{11} times the solar luminosity) and spiral arms with a luminosity that is smaller than that of the central part in spite of their much larger spatial area. The intensity ratio can change dramatically within some years or even some days. This indicates

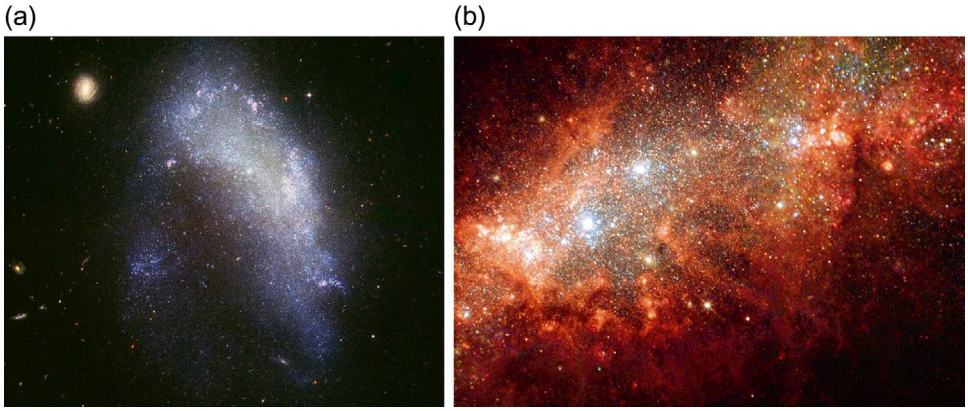


Fig. 6.40 Two Irregular galaxies are **a** NGC 1427A in the constellation Fornax (ESA, NASA) and **b** NGC 1569 in the constellation Camelopardalis (Hubble Space Telescope)

Fig. 6.41 Photograph of two colliding galaxies (NASA)

Table 6.1 Relative number of the different types among the brightest galaxies

Galaxy-type	E	S	L	I	Sum
Number	199	934	329	52	1514
%	13	61.1	21.5	3.4	99

special activities in the central region. The Seyfert galaxies are probably transition states in the development to normal spiral galaxies. Many Seyfert galaxies are strong X-ray sources. From their nonthermal energy distribution of the X-ray radiation, one concludes that the X-rays are emitted by electrons spiraling around magnetic field lines and matter falling into a huge black hole in the center of the galaxy. The spectra of Seyfert galaxies show broad emission lines emitted by regions close to the black hole, while sharp lines are found which might be emitted from regions far away from the black hole.

The core of Seyfert galaxies with a diameter of only a few pc can emit as much radiation power as 10^{11} stars of the galaxy. This can be seen in Fig. 6.42 where the core is much brighter than the other parts of the galaxy. The emitted spectra show some strong lines and a continuum with a nonthermal spectral distribution.

Quasars (see Chap. 5) were first discovered in 1960 aspoint-like intense radio sources. The observations showed that their luminosity fluctuated within a few days. This implies that their diameter must be smaller than a few light days (for comparison: The diameter of our own galaxy is about 10^5 light years).

Later on, quasars could also be identified as sources of visible radiation. However, even with the large 5 m telescope on Mount Palomar, they could not be distinguished from stars. Detailed observations finally showed that quasars are the active core of galaxies which contain at their center a large super-massive black hole. This black hole attracts all stars in its surroundings and finally swallows them. This is the source of the intense radiation, which appears in visible light as nearly point-like (such as a star) but also emits, in other spectral regions, huge amounts of energy.

Fig. 6.42 Seyfert Galaxy NGC 3147 in the constellation Dragon (ESA/HUBBLE & NASA, A. RIESS ET AL)

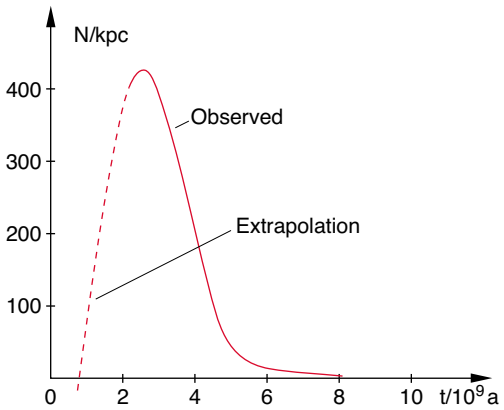


Fig. 6.44 Number of observed quasars per kpc distance interval is plotted as a function of the age of the universe at the time of their formation

Quasars belong to the brightest objects in the universe. Therefore, quasars with large distances from the earth can be still detected. In 2017, a quasar with a distance of 13 Billion light years (redshift $z = 7.6$) was detected, which must have been formed in the early age of the universe. It hosts a black hole with a mass of 1.3 billion solar masses.

The absolute magnitudes of quasars are between –25 ^{m} and –33 ^{m} (see Sect. 6.3x). This corresponds to luminosities of $\left(10^{12}-10^{14}\right)L_{\odot}$!!

In Fig. 6.44, the number of observed quasars per kpc distance interval is plotted as a function of the age of the universe at the time of their formation.

6.4.5 Galaxy Clusters and Super Cluster

The galaxies are not uniformly distributed in space but show clustering in certain parts of the universe. Between these clusters, large parts of space have a small density of galaxies. In such clusters, the density of galaxies can be 10^4 times that in the space between the clusters. Every deviation from a uniform mass density results in gravitational forces, which are directed toward the location of maximum mass density. These forces are called "**attractors**". The attractors influence the expansion of the universe which is only uniform on a large scale where the mean mass density is nearly homogeneous. Inside a galaxy cluster, the attractive

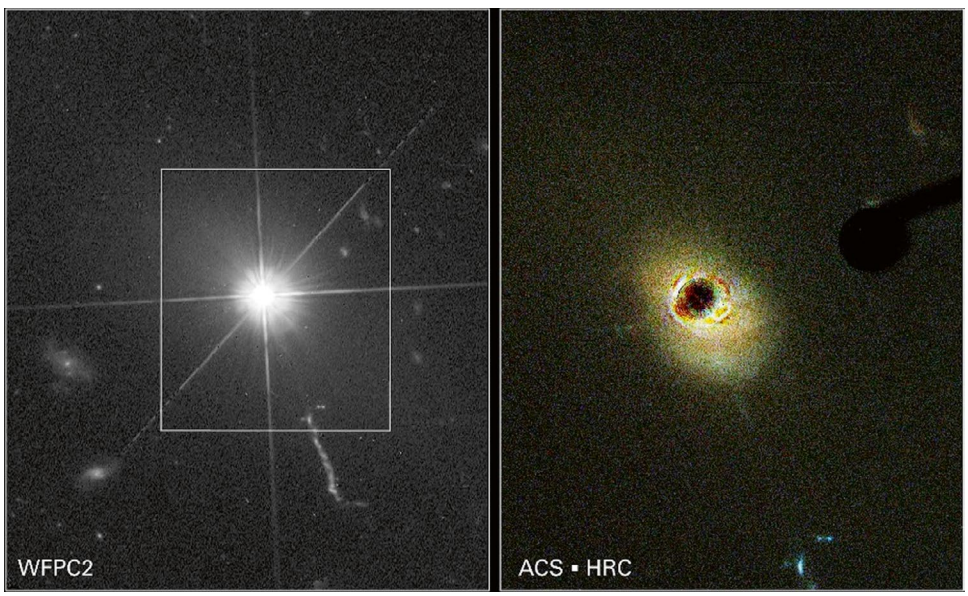


Fig. 6.43 Hubble images of quasar 3C 273. At right, a coronograph is used to block the quasar's light, making it easier to detect the surrounding host galaxy (NASA and J. Bahcall (IAS))

force between the galaxies of the cluster acts against the expansion. The same is true for the different galaxies which are held together by attractive gravitation. This means that the diameter of the cluster and also the diameter of a galaxy increases much less in time than the general expansion dR/dt of the scale parameter, while the distance between different clusters does increase as dR/dt.

The size of a cluster varies greatly. There are small clusters with diameters of 0.5–1 Mpc which contain only a small number of galaxies. They are often called galaxy groups. One example is our local group which contains 80 galaxies including our milky way, the Andromeda galaxy and a number of small galaxies. Four of these galaxies are of the spiral type, the others are elliptical and irregular types including some dwarf galaxies.

Most of the galaxy clusters contain several hundred to thousand galaxies. Their diameter ranges from 3 to 20 Mpc. One distinguishes between regular galaxy clusters, which have a nearly spherical shape and irregular galaxy clusters, which have no symmetric shapes and also no increasing galaxy density toward the center of the cluster. They have sometimes even several density centers and contain galaxies of all types.

Example

The majority of the galaxies in the Coma Cluster, which belongs to the regular galaxy clusters, are big elliptical galaxies and elliptical dwarf galaxies, while the Hercules galaxy cluster contains mostly spiral galaxies. ◀

Often in the central region of a galaxy cluster, massive elliptical galaxies are found, while the outer region of the cluster contains mainly spiral galaxies. Because of the large galaxy density and the corresponding large luminosity, galaxy clusters can be observed even at very large distances (up to several billion light years). In Table 6.2, some galaxy clusters are listed together with their characteristic data.

For sufficiently small distances between different galaxy clusters, they can interact by their gravitational attraction and form a system called a supercluster.

A Supercluster is a Gravitationally Bound System of Several Galaxy Clusters.

The spatial structure of superclusters has often either a filamentary form, or the clusters are arranged in a flat rotating disk. Between the different superclusters, bubble-like spatial regions are found which are nearly free of galaxies (voids).

Since the different galaxies in a cluster with sufficient large galaxy density are gravitationally bound (which can be proved experimentally), they move with special velocities (*peculiar velocities*) determined by the gravitational forces, i.e. by their total energy and their negative potential energy. This is completely analogous to the motion of stars in a galaxy (see Sect. 6.5.2). These local velocities superimpose the global escape velocity of the whole cluster.

The structure of very old clusters (large redshifts) is similar to that of younger nearby clusters (small redshifts). This indicates that the clusters are stable over a long period of time, which would not be the case without gravitational binding.

Measurements of the peculiar velocities allow the estimation of the mass of the total cluster. Comparing this value with that derived from the total luminosity of the cluster, it turns out that about 85% of the total mass does not emit electromagnetic radiation; it is a nonluminous matter. Since it cannot be seen, it is called **dark matter**. What this dark matter really is, i.e. which are the particles forming dark matter, is up to now completely unknown.

Besides the stars, the galaxy clusters contain intergalactic matter, which consists of hot gas (mainly ionized hydrogen (H^+), and dust. It often appears as a filamentary structure (Fig. 6.45).

The existence of this intergalactic matter is supported by measurements of X-rays, which were uniformly emitted from the area of the galaxy cluster. The spectral distribution of these X-rays corresponds to the thermal emission of matter at the temperature of $T = 10^7$–10^8 K. The spatial structure and the energy distribution are related to the cluster type. The regular clusters are stronger X-ray emitters than the irregular ones because the gas in the regular clusters has a higher temperature.

Table 6.2 Characteristic data of galaxy clusters

Name of cluster	Number of galaxies	Distance in 10^6 ly	Angular diameter in arc-minutes	Radial velocity (in 10^3 **km/s**)
Virgo	2500	70	700	1.2
Pegasus	100	220	60	3.8
Perseus	500	310	120	5.2
Coma Berenice	1000	350	180	6.7
Hercules	75	350		10.3
Ursa Majors	300	880	40	15.4
Leo	300	1000	35	20.0
Gemini	200	1250	30	21.6

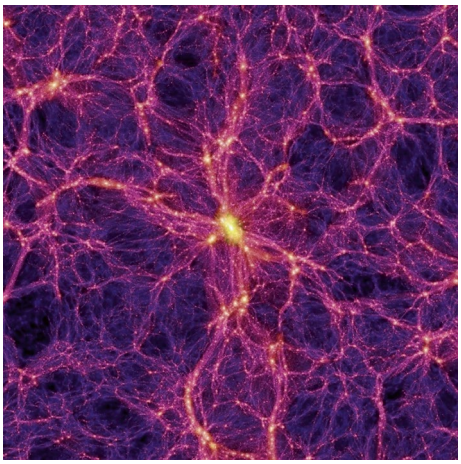


Fig. 6.45 The spatial filamentary structure of intergalactic matter. Result of mathematical modeling (Wikipedia)

6.4.6 Colliding Galaxies

Although the mean distance between different galaxies is large compared to their diameter, it happens during long time spans, that galaxies collide with each other. In the central regions of the clusters, the density of galaxies is high enough to realize collision probabilities of 1 collision per 100 million years. Since the lifetime of galaxies is much longer (up to 10 billion years), most galaxies suffer at least one collision during their lifetime [22].

The mean distance between stars is much larger than their diameter. Therefore, direct collisions between stars are very seldom, even if the galaxies collide. However, during such collisions, the velocities of the stars are generally changed, because the distances between stars change and therefore also the attractive forces, even at large collision parameters between the stars.

The intergalactic gas is much more affected by collisions between galaxies. Due to the strong changing forces during the collision, the gas can be even thrown out of the colliding galaxies [22].

During such a collision, the mass of the larger galaxy can attract the stars of the other galaxy. This causes an increase in the larger mass and a decrease in the smaller galaxy.

Spiral galaxies generally lose their spiral arms during a collision because the interstellar gas in the spiral arms, which represents the building material for the birth of stars, is attracted by the more massive central regions of the galaxies and is, therefore, missing in the regions of the spiral arms. This explains the fact that, in the central regions of regular galaxy clusters, mainly elliptical galaxies are found, while spiral galaxies are mainly present in the edge regions of galaxy clusters.

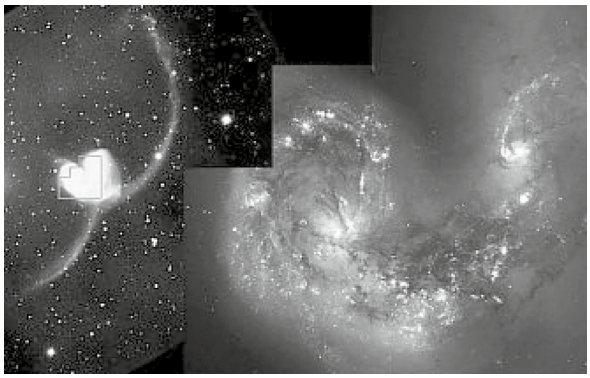


Fig. 6.46 The colliding antenna galaxies NGC 4038 and 4039 (NASA). The main figure is the enlargement of the insert on the upper left

Example

Examples of two colliding galaxies are the antenna galaxies NGC 4038 and 4039 in the constellations Corvus, which are shown in Fig. 6.46. Their name is due to the two brilliant gas jets that emerge from the colliding galaxies and which look like the antenna of an insect. The right part of the picture is the enlarged central intersection region of the two galaxies. The picture illustrates that, in this region, many new stars are born due to the turbulences of the gas- and dust clouds caused by the collision. ◀

6.4.7 Gravitational Lenses

According to Einstein's General Relativity Theory, photons are deflected when passing through a gravitational field G(r). The magnitude of the deflection angle Δß depends on the integral.

$$\Delta\beta = \int G(r)\mathrm{dr}0.$$

The first experimental confirmation of this proposition was performed by Sir A Eddington in 1919 at the sun eclipse, where the deviation of the light from a star nearly behind the sun was measured. The agreement between the measured value and the predicted value essentially contributed to the success of the General Relativity Theory and made Einstein worldwide famous.

If a massive object (a quasar or a massive galaxy) is located on the sightline between a distant radiating object and the earth the light which would not reach the earth without the massive object, is deflected and can reach the earth. The gravitational field of the massive object acts like a lens and focuses the light passing near the massive object onto the earth (Fig. 6.47).

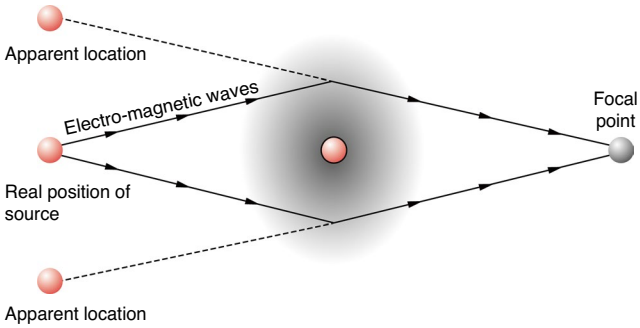


Fig. 6.47 Schematic picture of a gravitational lens

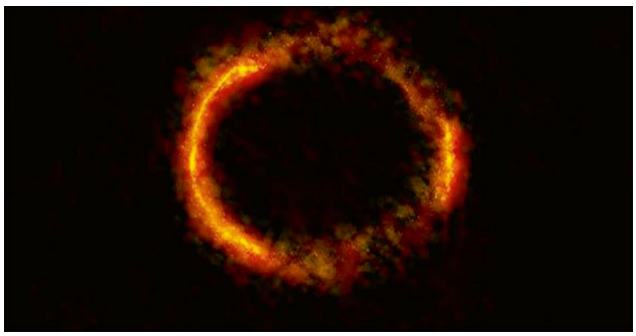


Fig. 6.48 Einstein ring of the galaxy 10 billion light years away, produced by the galaxy SDP 81, which is 4 billion ly away (ALMA NRAO/ESO/NAOJ), B.Saxton,NRAO/AUI/NSF)

If the mass distribution of the deflecting mass is symmetric and the emitting source, the deflecting mass and the earth lie exactly in the line of sight; the deflected light from the source appears on earth as a luminous ring (Einstein ring).

In Fig. 6.48, the Einstein ring of the light from a source at a distance of 10 billion light years is shown, which is imaged by a galaxy acting as a gravitational lens at a distance of 4 billion light years. This picture was taken by the ALMA (Atacama Large Millimeter/Submillimeter Array) of 66 radio antennas on the high plateau Chajnantor in the Atacama desert in Chile.

The strongest gravitational lenses are provided by galaxy clusters which represent the most massive objects in the universe. For non-spherical gravitational lenses, more than one image of the source can be produced. For example, the deflection by the spiral arms of galaxies results in two or more images of a point-like source. This is illustrated by Fig. 6.49, which shows the image of the quasar QSO2237+0305 (8 billion light years away) by a galaxy at a distance of 500 million light years. The image appears as a cross (Einstein Cross). Measurements of the splittings between the image points allow the determination of the mass of the gravitational lens.

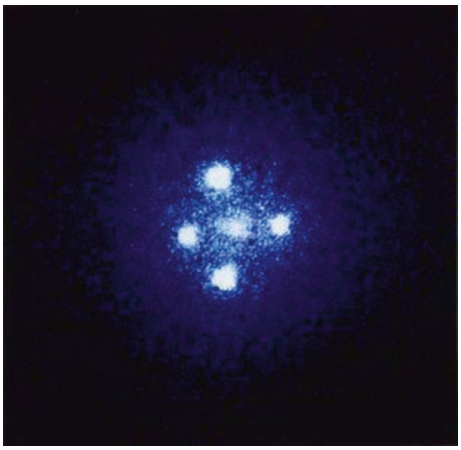


Fig. 6.49 Image of the quasar QSO2237+0305 (8 billion light years away) by a spiral galaxy at a distance of 500 million light years. The image appears as a cross (Einstein Cross)

The advantage of the gravitational lenses is the focusing of the radiation from far distant sources, which increases the light intensity, thus allowing the detection of radiation from very far objects, which could not be detected without this lens effect [23].

Summary

- The standard model of the universe is based on the assumption that the universe began at t = 0 as an extremely hot and dense fireball that expanded due to the high pressure and cooled down. The expansion continues up to now.
- Experimental proofs of the model are the cosmic background radiation with the spectral distribution of a thermal Planck radiation at the temperature $T = 2.7\,\mathrm{K}$; the abundance ratio of the elements H: D: He and the redshift of the radiation from far distant galaxies.
- Hints to the fact, that the universe, filled with stars, is finite, are given by Olbert's paradoxon.
- The universe can be regarded as a three-dimensional curved space with a scale factor R(t) (radius of curvature). The increase of R(t) with time corresponds to the expansion of the universe.
- The expansion started at t = 0 from a limited space with a scale factor $R(0) = 0$ (singularity). In the three-dimensional space, no point can be distinguished as the starting point of the expansion.
- During the early phase of the universe, the radiation energy density w exceeded the mass energy density ϱc^2 (radiation-dominated universe). While the mass density ρ decreased with $1/R^3$ the radiation energy density

w with $1/R^4$. At about 70.000 years after the big bang, the mass energy density became larger than the radiation energy density.
- The time progression of the scale factor R(t) depends on the mean mass density ϱ in the universe. For the critical mass density $\varrho = \varrho_c$, the total energy (kinetic energy of all particles + negative potential energy of the gravitational attraction) is zero.
- For $\varrho > \varrho_c$, the universe is closed, i.e. after having reached a maximum expansion value $R(t)_{max}$ it starts to collapse again. For $\varrho < \varrho_c$, the expansion never stops and $R(t) \to \infty$.
- Recent distance measurements of supernovae Ia in far galaxies have shown that the universe expands accelerated, in spite of the attracting gravitational forces. This is attributed to an up to now not understood dark energy.
- The age of the universe can be estimated from the measured Hubble constant and the assumed mass density. Most accurate results are obtained from the cosmic background radiation. The presently accepted value is $\tau = 13.8 \cdot 10^9$ ly.
- About 3–4 min after the big bang, the first elements H, D and He were formed. All heavier elements up to iron have been later produced by nuclear fusion in the central regions of stars. Elements above iron are produced by slow neutron capture during the final states of massive stars and by fast neutron capture in the hot envelopes of stars during a supernova explosion.
- After the recombination of electrons and protons about $5 \cdot 10^5$ years after the big bang, the radiation pressure becomes negligible and the neutral gas can clump together induced by spatial density inhomogeneities. The dark matter advances this process. As soon as the Jeans criterion is fulfilled, the gas cloud collapses and associations of stars are born.
- The Doppler shift of galaxies receding with the escape velocity v is really due to the expansion of the space, measured by the time derivative dR(t)/dt of the scale parameter R(t).
- The universe contains more than 10^{10} galaxies. Each galaxy consists of 10^9–10^{12} stars. There are different types of galaxies: Elliptical, spiral, lenticular and irregular galaxies. Our own galaxy belongs to the type Sb.

Problems

6.1 Show that at temperatures of $T = 10^9$ K and a baryon density $\varrho_B = 2 \cdot 10^{-3}$ kg/m^3 neutron capture by nuclei, according to.

$$^{A}_{Z}\mathrm{N} + \mathrm{n} \rightarrow ^{A+1}_{Z}\mathrm{N}$$

is more probable than the decay of free neutrons with a decay time of $\tau = 800$ s.

6.2 In the distribution of the thermal radiation of a black body, the wavelengths λ_0 are shifted to $\lambda > \lambda_0$ due to the expansion of the universe. Show that the thermal distribution is preserved but for a lower temperature. Calculate this temperature for an expansion velocity of 100 km/s.

6.3 For an adiabatic expansion of the universe with scale factor R(t), the change of the energy density with time is equal to the work necessary to expand the space with stars against the gravitational force.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\varrho \cdot c^2 \cdot R^3\right) = -p \cdot \frac{\mathrm{d}}{\mathrm{d}t}(R^3),$$

where p is the pressure exerted by the expanding matter. Derive from this equation the relation $w_m = \varrho \cdot c^2 \propto R^{-3} w_r = \propto R^{-4}$, where w_r is the energy density of the relativistic particles (photons and neutrinos).

6.4 We assume uniform spatial distribution of galaxies with the absolute magnitude M. Show that the ratio N_{m+1}/N_m of the numbers of galaxies with relative magnitudes m + 1 and m is equal to $10^{0,6} \approx 4$.

6.5 Derive the relations (6.83) and (6.84).

6.6 How large is a homogeneous density of stars with luminosity L the number of observable stars with apparent magnitude m within the solid observation angle $\Omega = 10°$?

References

1. K. Rudniki: Cosmogony of galaxies based on the spatial distribution of Matter. (Astron. Nachrichten 297, 317 (1976). Cosmological principle - Wikipedia
2. Astrophysics and Space Science Library book series (ASSL, volume 244)
3. Age of the universe - Wikipedia
4. https://www.nasa.gov/audience/forstudents/9-12/features/what-is-dark-matter.htm. Dark matter - Wikipedia Dark Energy, Dark Matter | Science Mission Directorate
5. Gonzalo J, Olmo: Open questions in Cosmology (Intech open 2012)
6. Cosmology - Wikipedia
7. Hetherington, Norriss S. (2014). Encyclopedia of Cosmology (Routledge Revivals): Historical, Philosophical, and Scientific Foundations of Modern Cosmology. Routledge. p. 116. ISBN 978–1–317–67766–6.
8. Dodelson, Scott (30 March 2003). Modern Cosmology. Elsevier. ISBN 978–0–08–051197–9.
9. Cosmic Journey: A History of Scientific Cosmology from the American Institute of Physics Cosmology and Theology (Stanford Encyclopedia of Philosophy)
10. https://lweb.cfa.harvard.edu/~dfabricant/huchra/hubble
11. https://www.nasa.gov/directorates/heo/scan/services/missions/universe/WMAP.htm

12. Curved Space - Special and General Relativity - The Physics of the Universe
13. Inflation (cosmology) - Wikipedia Guth, Alan H. (1997). The Inflationary Universe: The Quest for a New Theory of Cosmic Origins. Basic Books. pp. 233–234. ISBN 978-0201328400 Alan Guth - Wikipedia
14. 24583.pdf (bnl.gov) (PDF) Kaon Lifetime and Decay | Ilgaitis Prusis - Academia.edu
15. http://firstgalaxies.org/explore.htm How are early galaxies different than galaxies today? | Astronomy.com
16. https://arxiv.org/abs/1804.01318
17. The Standard Cosmological Model (caltech.edu)
18. Introductory Chapter: Standard Model of Cosmology | IntechOpen
19. H. Scheffler, H. Elsässer: Phyiscs of the Galaxy and Interstellar Matter (Springer, Heidelberg 1987)
20. Galaxy - Wikipedia Galaxies (nasa.gov)
21. https://en.wikipedia.org/wiki/Local_Group. The Universe within 5 million Light Years - The Local Group of Galaxies (atlasoftheuniverse.com)
22. NASA - When Galaxies Collide Galaxies: Collisions, Types and How They're Made | Space
23. https://en.wikipedia.org/wiki/Gravitational_lens An Introduction to Gravitational Lensing (thoughtco.com)
24. Local Group - Wikipedia

Our Home Galaxy: The Milky Way

7

Contents

7.1 Introduction . . . 215
7.2 Stellar Statistics and Star Populations . . . 215
7.3 Motions of the Stars in Our Neighbourhood . . . 217
7.4 The Differential Rotation of the Galactic Disc . . . 216
7.5 The Spiral Arms . . . 219
7.6 Globular Clusters . . . 220
7.7 Open Star Clusters . . . 222
7.8 The Center of Our Galaxy . . . 223
7.9 The Black Hole in the Center of Our Galaxy . . . 223
7.10 The Dynamics of Our Galaxy . . . 225
7.11 The Space Between the Stars; Interstellar Matter . . . 225
Summary . . . 229
Problems . . . 229
References . . . 229

7.1 Introduction

Already in the early seventeenth century *Galileo Galilei*, who observed for the first time the sky with a telescope, discovered that the Milky Way consists of an inconceivably large number of stars. Since these stars are much closer to us than stars in other galaxies, we can study them in more detail, in particular their spatial distribution, their motions and their spectral types [1].

7.2 Stellar Statistics and Star Populations

At the end of the eighteenth century *Friedrich Wilhelm Herschel*, one of the most important astronomers of his time, started a systematic counting of stars in different directions of the sky. He discovered that the star density is not isotropic but differs for the different directions. It showed a maximum in one direction and two minima in the two perpendicular directions. Such star counting was extended after the introduction of the Astro-photography and have found their perfection by the astro-satellite GAIA (see Sect. Xx) which measured the positions, distances and velocities of more than 10^7 stars in our galaxy with an astonishing accuracy.

For counting the real star densities in all directions, one has to take into account the absorption of the star light by gas- and dust clouds.

The principle of stellar statistics, which give the spatial distribution of stars with different luminosities, is based on the following consideration (Fig. 7.1):

We receive on earth from a star with luminosity L at the distance r_0 the energy flux density (= intensity)

$$I_0 = L/\left(4\pi r_0^2\right). \tag{7.1}$$

When we assume, that the density n(L) of stars with luminosity L is isotropically distributed, then all stars with luminosity L at a distance $r < r_0$ give on earth a larger intensity. Their number is

$$N_{\mathrm{L}}(I > I_0) = \frac{4}{3}\pi r_0^3 \cdot n(L). \tag{7.2}$$

Inserting r_0 from (7.1) gives

$$N_{\mathrm{L}}(I > I_0) = \frac{n(L) \cdot L^{3/2}}{3\sqrt{4\pi}} \cdot I_0^{-3/2}. \tag{7.3}$$

The total number of stars which give the intensity $I > I_0$, independent of their luminosity L is then

$$N(I > I_0) = \frac{I_0^{-3/2}}{3\sqrt{4\pi}} \int_{L=0}^{L_{\max}} n(L) \cdot L^{3/2} \mathrm{d}L. \tag{7.4}$$

Counting all stars which give a minimum intensity I_0 on earth allows the determination of the star density

$$N = \int n(L)\mathrm{d} \tag{7.5}$$

Such counting can now be performed in different directions of the sky Table 7.1 lists the result for stars with different apparent magnitudes m, which shows that the star density is not isotropically distributed. This indicates that we do

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_7

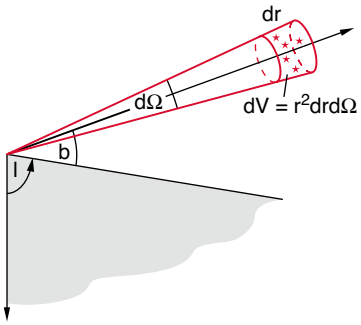


Fig. 7.1 Stellar Statistics: Number of Stars in the volume-element $dV = r^2 dr d\Omega$. The angle l gives the galactic length, b the galactic latitude

Table 7.1 Mean number of stars with apparent magnitudes m per square degree in two different directions with galactic latitude b, compared with the total number of stars on the sky

m	Mean number of stars per square degree		Total number on sky
	$b = 0{-}20°$	$b = 40{-}90°$	
9	2.28	0.90	62.090
10	6.18	2.33	165.600
11	16.48	5.80	$4.3 \cdot 10^5$
12	43.0	13.8	$1.1 \cdot 10^6$
13	110.2	31.5	$2.7 \cdot 10^6$
14	274.8	67.8	$6.5 \cdot 10^6$
16	1550	267	$2 \cdot 10^7$
18	7312	845	$1.4 \cdot 10^8$
20	28.190	2138	$5 \cdot 10^8$

not live in the center of our galaxy, but closer towards its edge. The detailed analysis of the data shows that the star distribution corresponds to a flat disc with a bulge in the central region (Fig. 7.2). Measurements in different spectral regions (visible, infrared, radio range) and the observation of neighbouring galaxies reveal that the flat disc consists of two major spiral arms and that many globular clusters form a nearly spherical halo around the center of our galaxy. The spiral arms contain a higher density of gas and dust particles. Here new stars are born. When we look onto other galaxies with their disc perpendicular to the line of sight, these new bright stars make the spiral arms appear very pronounced (Fig. 6.35). In Fig. 7.3 a picture of our home galaxy is drawn as constructed from all measurements performed up to now, in particular using data from the Spitzer satellite of the NASA.

It turns out that only two spiral arms exist, each emerging from one end of the central bar, contrary to the former assumption of 4 spiral arms.

The halo contains, besides the globular clusters, many old stars and gas clouds. From measurements of the differential rotation of the whole galaxy (See Sect. 7.5.3) it can be concluded that the halo must be filled mainly with dark matter of unknown origin.

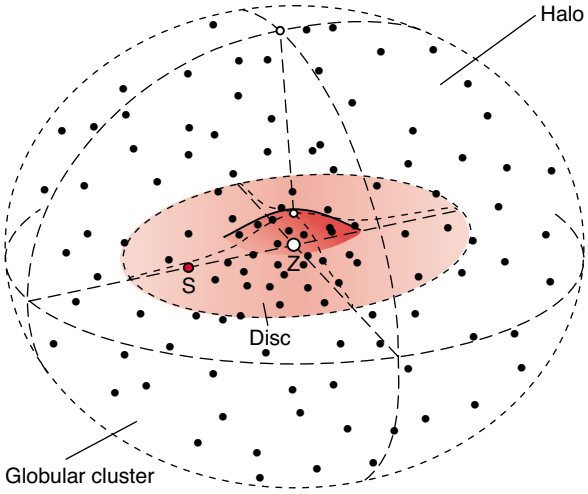


Fig. 7.2 Schematic representation of our galaxy, the Milky Way. The black points give the globular clusters in the halo

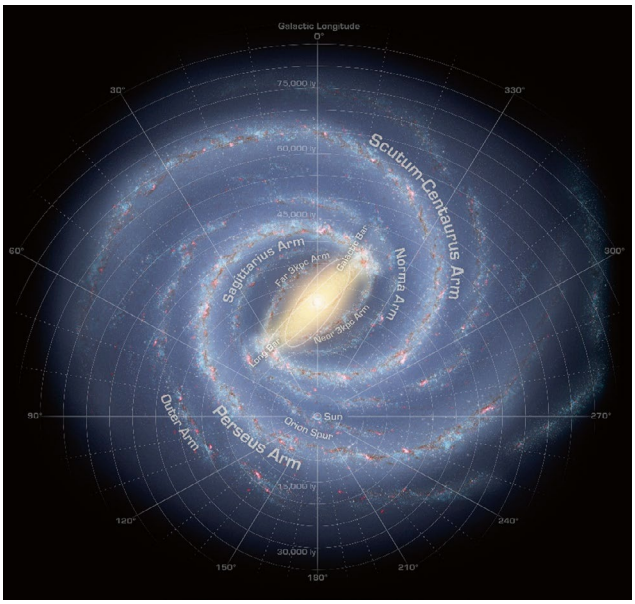


Fig. 7.3 Our Home galaxy, the Milky Way (NASA/JPL-Caltech/R.) Hurt (SSC/Caltech)

Harlow Shapley (1885–1947) was the first who could determine the distance of the globular clusters by measuring the period of variable Cepheids (see Sect. 5.x). Since the globular clusters are distributed within a sphere around the center of the galaxy, this also allowed the determination of the location of this center and estimations of the size of our galaxy.

The exact position of the earth within the galaxy and the characteristic distances are illustrated in Fig. 7.4, which also explains why the night sky is brighter and contains more stars in the northern summer time than in the winter time, because we look in a northern summer night towards the galactic center while in a winter night we look out of the galaxy.

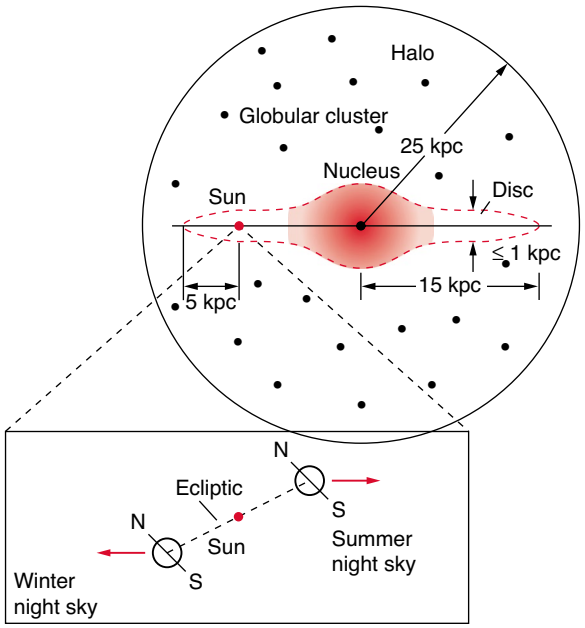


Fig. 7.4 Observation of our galaxy during summer and winter. The arrows give the observation directions in a summer night and a winter night

The sun is about 10 kpc away from the center. This corresponds to 2/3 of the radius of the galactic disc. As illustrated in Fig. 7.4 the white band of the milky way can be best seen from the northern hemisphere in summer nights, because then we look towards the center of the galaxy. But also during the longer winter nights the sky shows an impressive view of many brilliant stars because now we look into the Perseus spiral arm (Fig. 7.3), where many new stars are born.

Measurements of the spectral distribution of the star radiation show that in the galactic disc mainly young hot stars are found with a relative high percentage of heavy elements (heavier than helium). They are imbedded into the interstellar medium and rotate together with it around the center of the galaxy. They are classified as **population I**.

Different from this population I the stars in the globular clusters are old stars ($>10^{10}$ years), which contain mainly hydrogen and helium but only a small percentage of heavy elements. They move with the globular clusters on eccentric orbits around the galactic center mainly outside the galactic disc. They are named **population II**.

Between these two extreme classes there are many stars belonging to intermediate classes, as for instance our sun. These stars are called, together with the population I as **disc population**.

In Table 7.2 the relevant data of our galaxy are compiled.

7.3 Motions of the Stars in Our Neighbourhood

Although they were called in former times "*fixed stars*", stars are not staying at a fixed position but move with large velocities in different ways, as will be discussed in this section.

The velocity of a star relative to the observer A, can be partitioned into a radial component v_r and a tangential component v_t (Fig. 7.5). The radial component causes a Doppler-shift $\Delta\nu$ of the spectral line with frequency ν_0

$$\Delta\nu = \nu_0 \cdot v_r/c \tag{7.6}$$

Table 7.2 Characteristic data of our galaxy

General data	Our sun in our galaxy
Type: Spiral galaxy of type Sb	Stance of sun from center: 8.5 kpc
Age: $T = (10-12) \cdot 10^9 years$	Above the galactic plane: 14.5 pc
Diameter of disc: 30 kpc	Revolution velocity of sun around the galactic center: 220 km/s
Mean thickness: 2 kpc	Revolution period: $T = 2.4 \cdot 10^8 years$
Thickness at the location of our sun: 1 kpc	Escape velocity for $r = 8.5$ kpc: $v_e = 350$ km/s
Diameter of the central bulge: 10 kpc	
Thickness of the central bulge: 5 kpc	
Diameter of the halo: 50 kpc	
Total mass: $1.54 \cdot 10^{12} \cdot M_\odot$	
Mass of disc: $2 \cdot 10^{11} \cdot M_\odot$	
Mean density: $\varrho = 0.1\, M_\odot/\mathrm{pc}^3 = 7 \cdot 10^{-21} \mathrm{kg/m}^3$	
Absolute visual magnitude: $-20^m, 5$	
Mass portions in the disc	
Interstellar gas: 10%	
Interstellar dust: 0.1%	
Stars with $M > +3$mag: 80%	
Stars with $M < +3$mag: 10%	

Note Astrophysicists name all elements heavier than helium as metals. According to this nomenclature stars of the population I are metal-rich, those of population II are metal-poor

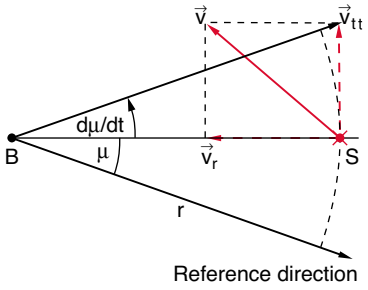


Fig. 7.5 Radial- and tangential components of the relative velocity of stars close to the sun as seen by the observer B

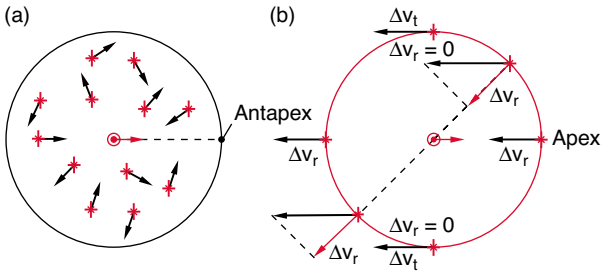


Fig. 7.6 **a** Statistical distribution of the velocities of stars including the sun in a reference frame that rotates with the galaxy. **b** Change of the radial velocities relative to the sun by the peculiar velocity of the sun directed towards the apex

The tangential component can be observed as the time change $\mu = \mathrm{d}\hat{e}_r/\mathrm{d}t$ of the direction vector $\hat{e}_r$ of the star with distance r, measured against a reference direction (for example the direction to a far distant star)

$$v_t = r \cdot \mu. \tag{7.7}$$

The angular part μ of the tangential component v_t is called "*proper motion*". For its determination the distance r of the star has to be known.

These relative motions measured from the moving earth, are converted into a local reference system called "local standard of rest" =LSR, which is defined as follows:

It is assumed that all stars within a volume that is small compared to the volume of the whole galaxy participate in the rotation of the galaxy in the same way, i.e. all these stars have the same rotational velocity about the center of the galaxy. Superimposed onto this rotational velocity are the individual **Peculiar-Velocities** of the individual stars, which can be determined from their velocities relative to the sun.

When measuring the radial velocities of the neighbouring stars relative to the sun, one observes, that they are not isotropically distributed, as one would expect for randomly distributed star motions (Fig. 7.6a).

The reason for this result is the motion of the sun, relative to these stars. The point on the celestial sphere towards which the sun is moving, is called the **Apex**, the point on the opposite side is the **Antapex** (Fig. 7.6b).

Therefore, the relative velocities in the direction of the apex are smaller by the velocity of the sun, while stars with velocities perpendicular to the line apex-antapex do not suffer changes of their radial velocities but only of their proper motion.

The local reference system is now defined in such a way that the vector sum of all star velocities relative to the sun is just equal to the opposite sun velocity. Then the mean value of the vector sum of all peculiar velocities (including that of the sun) is in this local reference system just zero.

The velocity of the sun in this reference system is

$$v_\odot = 19.7 \text{ km/s}$$

In order to define the direction of the sun velocity one has to define the Apex-coordinates, which depend on the coordinate system. In the galactic coordinate system (see Sect. 1.3) the sun coordinates are: $l = 56°$, $b = +23°$ (Fig. 1.6). In the equator-system they are $\alpha = 18^h\,0^m, \delta = +30°$.

The apex lies in the constellation *Hercules*.

Note: Nowadays the mainly used reference system is a system of selected reference galaxies.

7.4 The Differential Rotation of the Galactic Disc

The random peculiar velocities of the stars, discussed in the previous section, are small compared to the velocity of the common rotation of all stars and gas clouds around the galactic center. Although one can deduce already from the oblateness of the galactic disc that the disc must rotate, it was not before 1930 that the magnitude of the rotation could be quantitatively measured and the angular velocity $\omega(r)$ determined. The measurements were based on a method explained in Fig. 7.7.

We assume that all stars move on circular orbits around the galactic center. We select a star St with the distance r from the sun and the distance R from the galactic center, and the galactic angle l between the line St-sun and the line from the sun to the galactic center C. The velocity vector of its rotation around the galactic center is the tangent to the circle with radius R (red line). It has the two components v_r and v_t with respect to the line star-sun. The velocities of star and sun $v_{St} = \omega \cdot R$ and $v_\odot = \omega_0 R_0$ can be expressed by the angular velocity and the distance R resp R_0 to the galactic center. The sine theorem for the triangle C-St-S_0 gives

$$\frac{\sin l}{(\sin(90° + \alpha))} = \frac{R}{R_0}$$

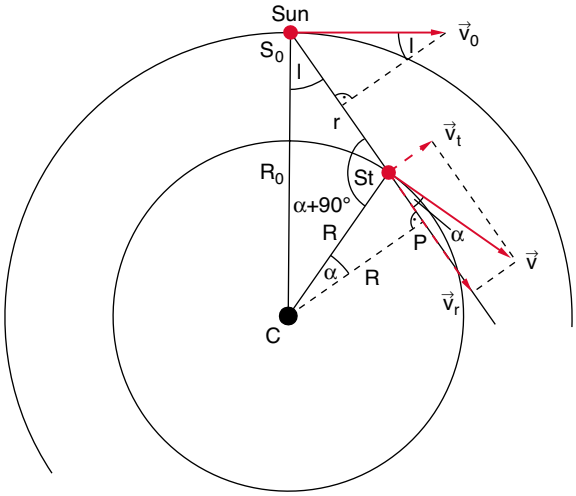


Fig. 7.7 Illustration of the derivation of the Oort's equations

$$\Rightarrow \cos\alpha = \frac{R_0}{R} \cdot \sin l. \tag{7.8}$$

The observed radial velocity of the star relative to the sun is then

$$\begin{aligned} v_r &= v \cdot \cos\alpha - v_0 \cdot \sin l \\ &= \omega \cdot R \cdot (R_0/R) \cdot \sin l - \omega_0 R_0 \cdot \sin l \\ &= R_0(\omega - \omega_0) \cdot \sin l. \end{aligned} \tag{7.9}$$

Its tangential velocity relative to the sun is according to Fig. 7.7

$$v_t = v \cdot \sin\alpha - v_0 \cdot \cos l. \tag{7.10}$$

From the rectangular triangle C-P-S_0 we obtain the relation

$$R \cdot \sin\alpha = R_0 \cos l - r, \tag{7.11}$$

From (7.10) we finally get

$$v_t = R_0(\omega - \omega_0) \cdot \cos l - \omega \cdot r. \tag{7.12}$$

When v_t, v_r r, R_0 and l are measured, the angular velocities ω and ω_0 can be determined from the two Eqs. (7.9) and (7.12). These measurements confirm, that the angular velocity ω(R) of the galactic rotation depends on the distance R from the galactic center.

Our galaxy does not rotate like a rigid body but shows a differential rotation ω(R). l.

Figure 7.8 illustrates how this differential rotation affects the star velocities relative to the sun: If ω(R) decreases with increasing R, stars with $R<R_0$ have relative velocities in the clockwise direction, while for stars with $R>R_0$ their relative velocities are anti-clockwise. Here we have assumed that the peculiar velocities are small compared to the rotational velocities.

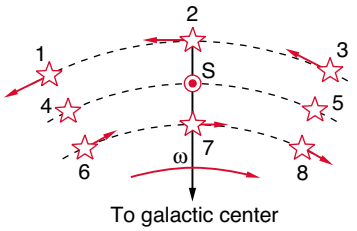


Fig. 7.8 Rotational velocities of stars close to the sun with different distances R_i from the galactic center measured relative to the sun

For stars close to the sun ($r \ll R_0$) the difference $\omega - \omega_0$ is small. We can therefore neglect in the Taylor expansion

$$\omega - \omega_0 = \left(\frac{d\omega}{dR}\right)_{R_0} \cdot (R - R_0) + \cdots \tag{7.13}$$

the higher order terms. With $\omega = v(R)/R$ we obtain

$$\left(\frac{d\omega}{dR}\right)_{R_0} = \frac{1}{R_0}\frac{dv}{dR} - \frac{v_0}{R_0^2}. \tag{7.14}$$

Inserting into (7.13) gives

$$\omega - \omega_0 \approx \frac{1}{R_0^2}\left[R_0\left(\frac{dv}{dR}\right)_{R_0} - v_0\right](R - R_0). \tag{7.15}$$

For $R_0 \gg r$ we can use the approximation $(R - R_0) \approx r \cdot \cos l$. The exact formula (7.9) for the relative radial velocity can then be approximated by

$$\begin{aligned} v_r &\approx \left[\frac{v_0}{R_0} - \left(\frac{dv}{dR}\right)_{R_0}\right] r \cdot \cos l \cdot \sin l \\ &\approx A \cdot r \cdot \sin 2l \end{aligned} \tag{7.16}$$

The constant

$$A = \frac{1}{2}\left[\frac{v_0}{R_0} - \left(\frac{dv}{dR}\right)_{R_0}\right] \approx 14 \frac{\text{km}}{\text{s} \cdot \text{kpc}} \tag{7.17}$$

is named **1. Oort's Constant** after the Dutch astronomer *Jan Hendrik Oort* (1900–1996) who has calculated this constant 1927, based on measurements of the radial velocities and peculiar velocities of many stars.

In an analogue way the relative tangential velocity (7.12) in the local reference system can be derived as (see Problem 7.2)

$$v_t = \left[\frac{v_0}{R_0} - \left(\frac{dv}{dR}\right)_{R_0}\right] r \cdot \cos^2 l - \omega_0 \cdot r \tag{7.18}$$

With $2\cos^2 l = 1 + \cos 2l$ this can be written as

$$v_t = A \cdot r \cdot \cos 2l + B \tag{7.19}$$

where the 2. Oort's constant B is

$$B = -\frac{1}{2}\left[\frac{v_0}{R_0} + \left(\frac{dv}{dR}\right)_{R_0}\right] \approx -12\frac{\text{km}}{\text{s}\cdot\text{kpc}}. \tag{7.20}$$

From (7.18) and (7.20) we get

$$A - B = \frac{v_0}{R_0} = \omega_0,$$

$$A + B = -\left(\frac{dv}{dR}\right)_{R_0} = R_0\left(\frac{d\omega}{dR}\right)_{R_0}. \tag{7.21}$$

The determination of the two Oort's constants allows to obtain the angular velocity ω_0 of the galactic rotation at the distance R_0 from the galactic center as well as its derivative $d\omega/dR$ R.

In order to achieve this, one has to measure the radial velocity v_r as well as the tangential velocity v_t of many stars at different distances R from the galactic center.

Note: These relative velocities due to the galactic rotation are superimposed on the peculiar velocities of the individual stars. Also the average of the vector sum of the peculiar velocities of all stars in a certain volume is zero in the local reference system, although each star has a peculiar velocity $v \neq 0$. These peculiar velocities can have also components perpendicular to the galactic disc. However, the displacements z out the plane $z = 0$ are small because the gravitational attraction by the whole matter in the disc brings the escaping stars back into the disc, The path of such stars, which appears as an oscillation in z-direction and as a standing wave within the rotating disc, is schematically indicated in Fig. 7.9. The thickness of the galactic disc is small (the thickness at the location of the sun is 1 kpc compared with a disc diameter of 30 kpc).

The approximations (7.21) are restricted to the range with $r \ll R_0$ in the surroundings of the sun. For measurements of the rotation velocity v(R) one needs radio astronomy, because the visible radiation is in several directions absorbed by gas-and dust clouds. The measuring technique

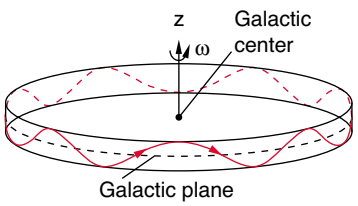


Fig. 7.9 Schematic representation of the orbit of a star with a velocity component v_z vertical to the central plane of the galactic disc

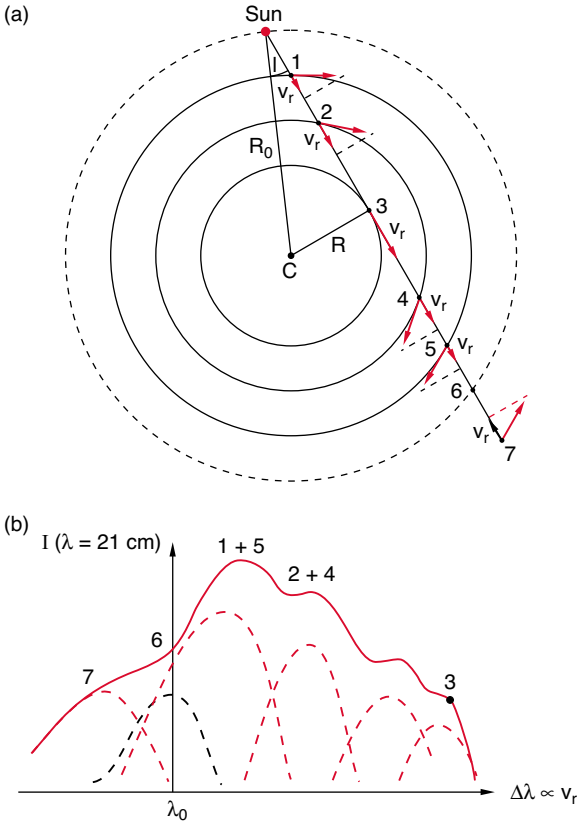


Fig. 7.10 Principle of the measurement of radial velocities relative to the local reference system based on the spectral width of the 21 cm line

is illustrated in Fig. 7.10. The radial velocity of hydrogen molecular clouds (H_2) and of atomic hydrogen (H) is measured along the line of sight for different radii R. Because of the Doppler-shift this gives a spectral profile of the H_2 rotational lines resp. the 21 cm line of atomic hydrogen. This measured line profile represents the superposition of many Doppler-shifted lines from atoms or molecules at the different radii R along the line of sight. The deconvolution yields the different parts I(v,R), which reflect the rotational velocities v(R) (Fig. 7.10b).

The results of these and many further measurements yield the rotation curve v(R) of the galactic disc (Fig. 7.11). The points represent the measured rotational velocities of gas clouds. The red curve was calculated by the Dutch-American astronomer *Marten Schmidt*, based on a special model of the galaxy.

If the total mass of the galaxy were restricted to a volume with a radius $R < 20$ kpc, a Kepler curve would be expected for $R > 20$ kpc, where $v(R) \propto R^{-3/2}$ The fact, that the curve ω(R) still rises for distances $R < 20$ kpc, indicates that there must be still mass for $R > 20$ kpc surrounding the galaxy in a halo. Since this mass has not been seen by emitted electro-magnetic radiation, it is called "dark matter".

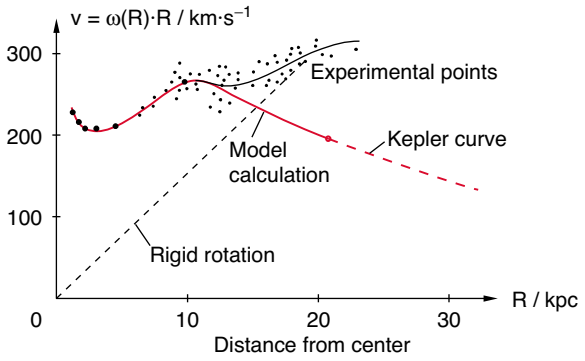


Fig. 7.11 Measured values (black points) of the rotational velocities of stars as the function of the distance R from the galactic center and comparison with Kepler's laws and with a rigid rotator

7.5 The Spiral Arms

The question whether our galaxy has a spiral structure like many other galaxies, and which experimental methods are best to find out about this, has been discussed for a long time. Only the combination of optical and radio-frequency measurements has brought sufficient accurate results to construct a map of our galaxy with its spiral structure (Fig. 7.3). The method of the measurements of the rotation curve and the determination of the distances of stars discussed in the previous sections has brought the essential insight into the structure of our galaxy.

Measurements of the Doppler shift (for example of the 21 cm line in radio-astronomy, which corresponds to the hyperfine transition $F = 2 \rightarrow F = 1$ in the $1^2S_{1/2}$ ground state of the H-atom) along the line of sight from the earth to the observed object, the velocity component of the H-atoms can be determined. This allows to obtain the Oort's constants A and B and therefore according to (7.21) also the angular velocity $\omega(R)$ of the differential rotation of the galaxy. Both informations together give the distance r of the emitting H-atoms from the sun. The H-atom clouds represent the major component of the interstellar gas. As shown in Fig. 7.10 the method is not completely unambiguous, because stars with $R < R_0$, which are closer to the galactic center than the sun, give for two different distances the same Doppler-shift. However, this problem can be solved because tuning the angle l of the galactic length (i.e. turning the line of sight) the farther away source will show a smaller tuning range of l than the closer component.

The intensity of the 21 cm radiation is proportional to the density of the H-atoms. One therefore receives maximum signals from regions with a high density of H-atoms. We have seen in Sect. 5.x that the Jeans Criterion for the contraction of a gas cloud is most probably met in regions with a high gas density. Therefore, these are the regions where new stars are born.

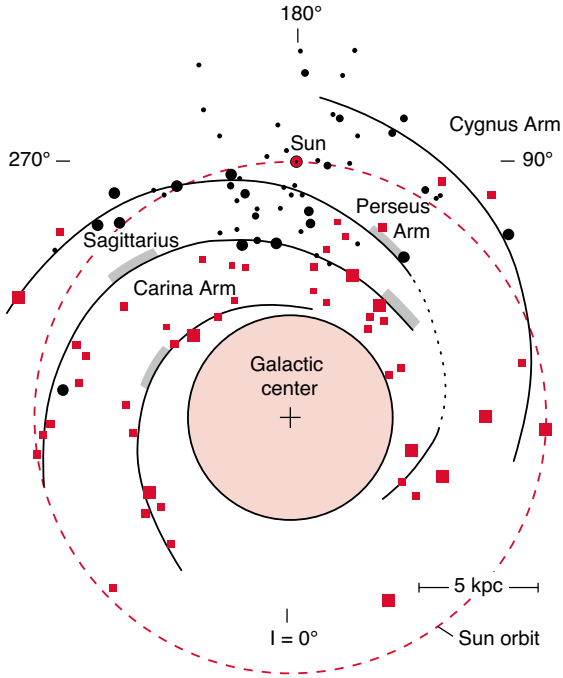


Fig. 7.12 Former assumed spiral structure of our galaxy with 4 spiral arms, based on optical (black points) and radio radiation measurements (red squares) (from H. Karttunen et al. Fundamental Astronomy 3rd ed. Springer, Heidelberg 1994)

One can observe in the spiral arms of other spiral galaxies bright new stars of spectral type OB (see Sect. 5.2) and extended clouds containing H II (ionized hydrogen atoms) neutral H-atoms, H_2-molecules and dust particles. Therefore, one can assume that in our galaxy similar conditions are present.

Although the major part of the luminosity of the spiral arms originates from these bright stars, the main part of the mass is contributed by the dust clouds.

The first attempts to find out about the spiral structure of our galaxy was performed 1952 by *Oort, Westerhouse* and *Kerr* as well as by *Morgan*. These authors tried a combined analysis of the results obtained from measurements of the 21 cm line by the radio telescopes in Australia and Holland. *Y. M.* and *Y. P Georgelin* determined 1976 the distances of the H II regions by a combination of radio- and visible measurements. More detailed pictures are obtained by the radio-radiation of the CO-molecules at $\lambda = 0.07$cm, measured at the Australian radio-telescope. The analysis of these measurements are shown in Fig. 7.12 where 4 spiral arms might be recognized, named after the star constellation which have the same sight of view from the earth as the main part of the spiral arms, namely Cygnus arm, Perseus arm, Orion arm (also named local arm where our solar system is located) and Carina arm.

Later, more accurate measurements showed, however, that there are really only 2 spiral arms (Figs. 7.3 and 7.14), which emerge from opposite sides of the central bar.

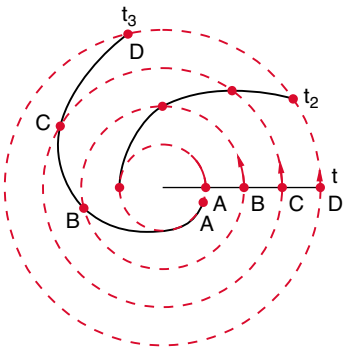


Fig. 7.13 Illustration of the formation of the spiral structure

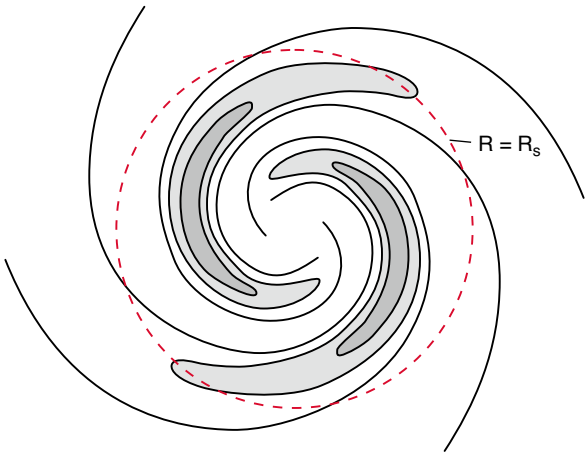


Fig. 7.14 Spatial density distribution of the rotating galactic disc calculated according to the density wave theory

The question is now, how the spiral arms are created?

In Fig. 7.13 four gas cloud s A, B, C, D at different distances from the galactic center are shown. Because of the differential rotation of the galactic disc the angular velocity $\omega(R)$ is larger for the cloud A than for B, C or D. If all four clouds are at the time t_1 located in a straight radial line, they will wind up into a spiral form in the course of time. Since the inner cloud A needs less than 10^8 years for a whole revolution, each spiral arm should develop 100 windings during the lifetime $\tau = 10^{10}$ years of the galaxy. This contradicts the observations which show that our spiral galaxy has only 2 arms.

In order to solve this problem, *Bertil Lindblad* proposed a density-wave- hypothesis which was further elaborated 1963 by C. C. Lin and F. H. Shu.

This density wave theory is based on the assumption, that, caused by a non-axial symmetric gravitational perturbation at the distance R_s a density wave moves on an elliptical orbit around the galactic center, which, according to the calculations, really results in a gas compression in spiral arms, which moves with the constant rotational velocity $\omega(R_S)$ around the galactic center. The essential point is now, that the particles can have a rotational velocity, that is different from that of the density wave. This implies that in the maximum of the density wave not always the same particles are present. They move for $\omega(R) > \omega(R_S)$ from behind the density wave into its maximum and overtake it, while for $\omega(R) < \omega(R_S)$ the maximum of the density wave overtakes the particles.

The analogon to this situation is a traffic jam on the motorway, caused by a slowly moving truck convoy. The maximum of the car density moves with the speed of the truck convoy, while always new cars reach the traffic jam from behind and leave it after having passed the maximum. In Fig. 7.14 the results of these calculations are illustrated. On the red dashed circle ($R = R_S$) particle velocity and the velocity of the density maximum are equal.

According to the calculations such a density wave structure can be stable during the whole age of the galaxy.

In the density maxima stars can be formed, if the density exceeds the critical Jeans density for the contraction of gas clouds. The larger the mass of the contracting cloud is, the more stars are created. This explains, why in the spiral arms with a larger density so many young stars are found.

An alternative theory explaining the creation of the spiral arms assumes that gravitational perturbations caused by closely passing neighbour galaxies (e.g. the Magellan's Clouds) cause tide effects resulting in density fluctuations. A known example of gravitational perturbations by colliding galaxies are the *Antenna galaxies* (Fig. 6.x) or the Whirlpool Galaxy M 51.

7.6 Globular Clusters

Globular Clusters are nearly spherical accumulations of 10^4–10^7 stars (Fig. 7.15) which move around the galactic center. The mean diameter of their orbits is below 30 kpc. Their large mass indicates that they have been formed by the collapse of very large gas clouds of low gas density. (See Sect. 6.z) Within the halo around our galaxy about 150 globular clusters have been observed (Fig. 7.2). Around elliptical galaxies often several thousand globular clusters exist. The stars in these globular clusters belong to the oldest stars in our galaxy. They are about 10–15 billion years old and must have been formed before the disc of our galaxy [2]. Their age can be inferred from their colour-luminosity diagram (Fig. 6.xx). Such a diagram is shown in Fig. 7.16 for the globular cluster M3 in the constellation Hunting dogs (Canes Venatici).

Instead of the surface temperature of a star often the colour index $CI = m(\lambda_1) - m(\lambda_2)$, which describes the brightness difference between two wavelength ranges. Astronomers have agreed upon an international color

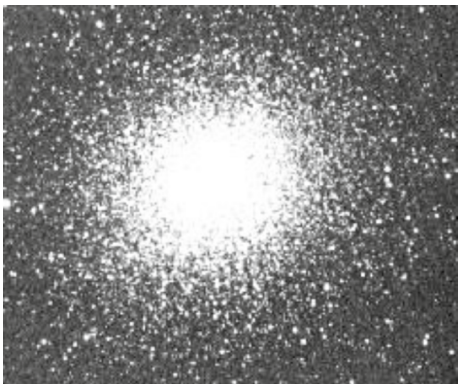


Fig. 7.15 Globular Cluster M13 (NGC 5272) in the constellation Canes Venatici (hunting dogs) at a distance of 34.000 ly containing about 0.5 million stars

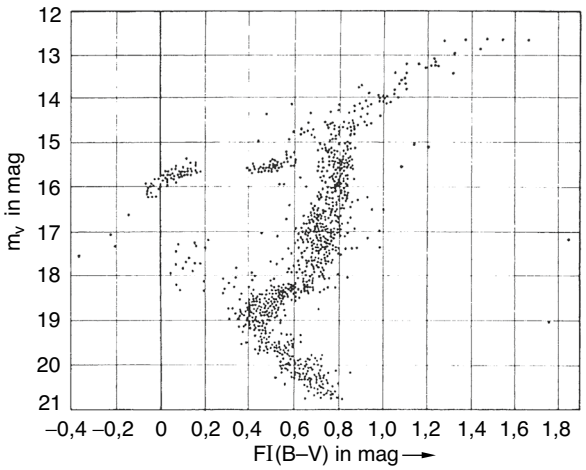


Fig. 7.16 Colour-Luminosity diagram of the globular cluster M3. The absolute luminosity L is plotted against the colour index (from F. Gondolatsch, S. Steinacker, O. Zimmermann: Astronomie (Klett-Verlag, Stuttgart 1980)

system which corresponds to the transmission curves of special color filters, which are marked by the letters U (ultraviolet), B (blue), V (visible), and I (infrared).The filtered transmission ranges have a typical spectral width of 100 nm.

On the lower part of the main sequence in Fig. 7.16 are the stars with low masses ($M < 0.06 M_\odot$) which have long lifetimes because of their low mass (see Sect. 6.yy) Since a large fraction of the stars has already left the main sequence (upper right above the kink in Fig. 7.16) and has converted into red giants, the cluster must be older than the mean time on the main sequence. The age determination is based on the position of the kink in the colour-luminosity diagram, where the stars are just leaving the main sequence. Model calculations of star developments give information about the time, the stars spend on the main sequence. Such calculations give an age of 8–12 Billion years, depending on the model and its calibration.

Detailed investigations of globular clusters provide us with important information about the development of stars (see Sect. 6.xy). Since all stars in the cluster have been born at the same time but have different masses, their position in the HR-diagram tells us about the difference in the development of stars with different masses. Since all stars in the cluster have the same distance r from the earth (the diameter 2R of the cluster is very small compared to its distance r), the ratio of their measured apparent luminosities gives also the ratio of their absolute luminosities. Therefore, their classification into the HRD is much more accurate than for stars with different distances which can be determined only with certain error bars.

In former times the age of the globular clusters had been estimated as $T > 10^{10}$ years. However, the much more accurate distance measurements by the astrometric satellites HIPPARCOS and GAIA have proved that their distance is larger than originally assumed. Therefore, the absolute luminosity of the stars in the cluster is larger, which means that their lifetime is smaller ($\approx 8 \cdot 10^9$years). This solves an old discrepancy, because the older distance measurements gave an age of the globular clusters which was larger than the age of the universe obtained from the Hubble constant.

The clusters also contain fast pulsating stars (see Sect. 6.yx), the RR-Lyra stars, with pulse periods between 0.2 and 1.2 days. From their period and their luminosity their distance can be determined. Since the globular clusters are uniformly distributed in a spherical halo around the center of the galaxy, this allows the determination of the position of the galactic center.

The globular clusters orbit around the galactic center with revolution periods of several million years. Their orbits are randomly distributed, they can be also vertical to the disc plane. During their revolution they must pass through the galactic disc, if their revolution radius is smaller than the disc radius. Their density decreases with increasing distance from the galactic center (Table 7.3).

Table 7.3 Spatial density of globular clusters as a function of the distance R from the galactic center

R/kpc	Number/(10 kpc)^2
0–2	270
2–4	47
4–6	30
6–8	6
8–10	5
10–15	1
15–20	0.3
20–30	0.1
30–40	0.02

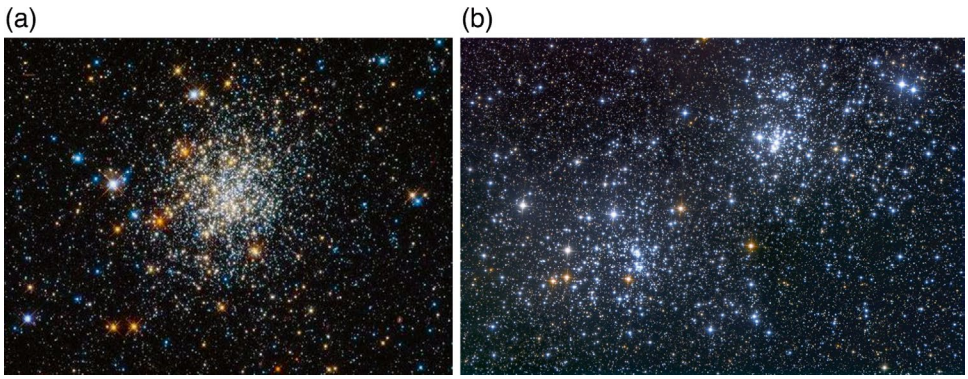


Fig. 7.17 **a** Open Star Cluster NGC 411 in the small Magelanean Cloud (Hubble Space Telescope NASA). **b** Double star cluster h + chi in the constellation Perseus (Hubble Space Telescope NASA)

7.7 Open Star Clusters

Open star clusters are star accumulations of irregular shape with about 10–1000 stars which have been born all at the same time by the contraction of gas and dust clouds. Contrary to globular clusters the star density is here much smaller. They are concentrated in the galactic disc and in the spiral arms. It is estimated that their number in our milky way exceeds 10.000. However, most of them cannot be seen from the earth because they are hidden by absorbing clouds in the disc.

Famous examples are the Plejades (Fig. 5.22), the Hyades or the double star cluster h + chi in the constellation Perseus (Fig. 7.17b).

A possible classification of open star clusters is based on the Hertzsprung-Russel diagram. Since all stars of the cluster have been born at the same time, their position in the HRD depends on their mass. For young clusters most stars are still on the main sequence. Only the most massive stars have left the main sequence and have drifted to the upper right where the red giants are positioned. Therefore, there appears in the HRD a kink, where the transition to the red giant starts. Often the equivalent colour-luminosity diagram is used instead of the HRD (Fig. 7.18). Here the apparent luminosity of stars is plotted against the colour-index. The colour index is a measure for the difference of the apparent luminosities in different wavelength ranges which are defined by the transmission regions of selected spectral filters (Fig. 7.19).

The open clusters offer a good possibility to determine the mass of the stars in the cluster and their absolute luminosities. Since the diameter of the cluster is very small compared to its distance from the earth, all stars in the cluster have approximately the same distance from us. The measured relative apparent luminosities are therefore also the relative absolute luminosities. It is therefore possible to construct the colour-luminosity diagram directly from the experimental data with high accuracy. If the distance to the cluster is known (after the very accurate measurements

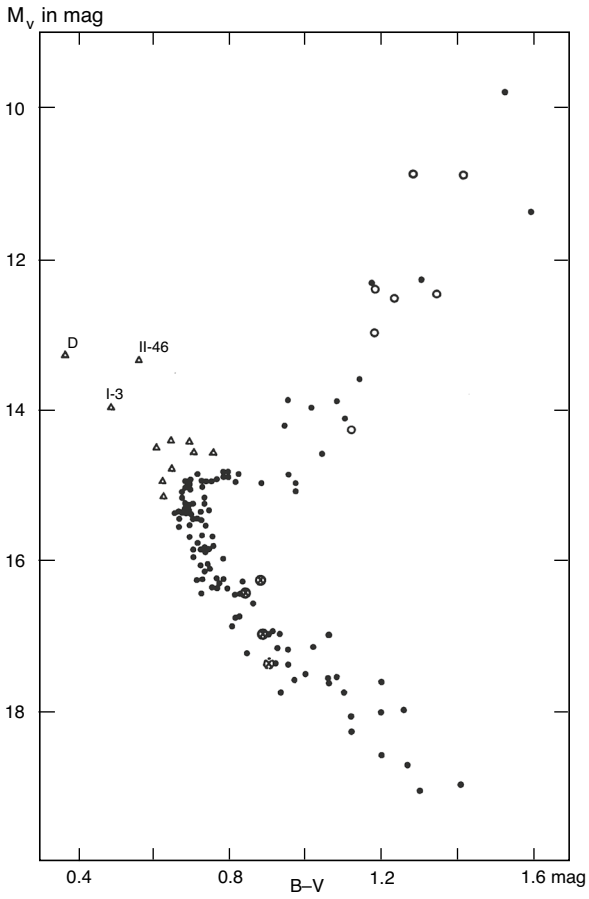


Fig. 7.18 Colour-Luminosity diagram. The absolute magnitude M is plotted against the colour index (the difference of the apparent magnitudes (m_b–m_v) in the blue spectral range and the whole visible range for the oldest open cluster NGC188

of the astrometric satellite GAIA this is the case for most clusters in our and in neighbouring galaxies). the position of the kink gives the age of the cluster. For young clusters a distinct gap appears between the Main Sequence and the region of the red giants, which gradually disappears with increasing age of the cluster. For instance, it turns out that the age of the Pleiades is only some 10^7 years, while the

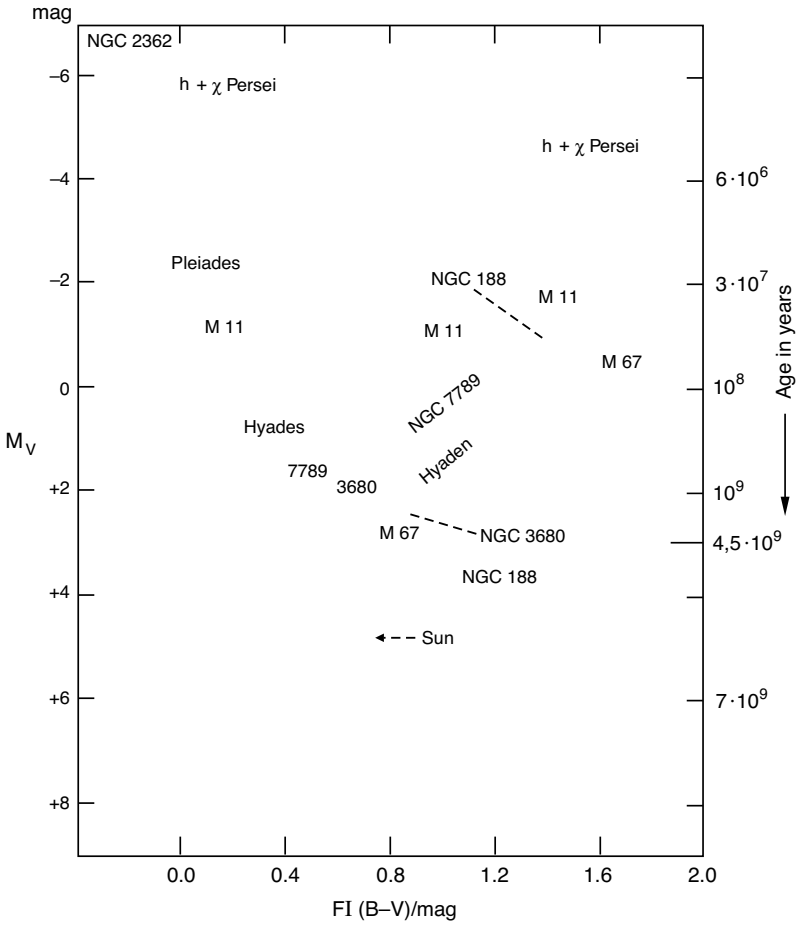


Fig. 7.19 Colour-luminosity diagram for galactic clusters with different ages. The youngest clusters NGC2362 and h + Persei are only some million years old, while the age of the oldest cluster NGC188 is (5–8)·10^9 years (after Sandage et al.: Publ. Astron. Soc. Pac. 68, 498 (1956)

Hyades are about 10 time older. The oldest open star clusters (e.g. M67 in the constellation cancer) approach already the age of the globular clusters with some 10^9 years.

7.8 The Center of Our Galaxy

The bulge of the disc around the galactic center is similar to a small elliptic galaxy. Seen from the earth the galactic center lies in the constellation Sagittarius. The star density increases rapidly with decreasing distance from the center (Table 7.4). Here the oldest stars of the disc population can be found, which had been formed relatively early at the formation of the galaxy.

Table 7.4 Spatial star density in the galactic center in the unit $M_\odot/pc^3$ as a function of the distance r from the center

r/pc	$\rho/M_\odot \cdot \mathrm{pc}^{-3}$
1	$4 \cdot 10^5$
10	$7 \cdot 10^3$
100	$1 \cdot 10^2$
1000	5

These stars cannot be measured with optical techniques because visible light is completely absorbed by the interstellar matter. During the last years investigations of the galactic center have been performed from satellites with infrared and X-ray radiation. Using the large telescope at the Paranal in Chile, and the 10 m Keck telescope on Hawaii, detailed measurements with infrared light have brought an astonishingly good spatial resolution of the galactic center, which allowed the spatial resolution of individual stars in spite of the extremely high star density (Fig. 7.20). Not only the position but also the velocities v(r) of the stars at the distance r from the center could be measured from the Doppler-shift of spectral lines. If r and v(r) are known, the total mass inside the orbit of a star with distance r from the center can be determined using Kepler's 1. Law.

7.9 The Black Hole in the Center of Our Galaxy

Theoretical Astrophysicists have assumed since long that in the center of galaxies very massive black holes should exist [2]. This assumption was experimentally confirmed by recent

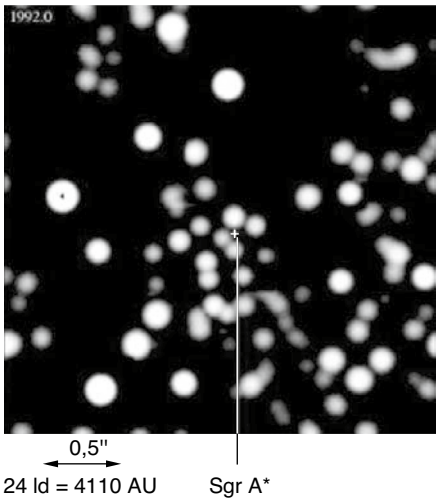


Fig. 7.20 Infrared observation of star positions and velocities in the center of our galaxy, measured with very high spatial resolution over a period of 8 years (R. Genzel, Max-Planck Institute of Extra-Terrestrial Physics, Garching Germany)

Fig. 7.21 The center of our milky way (ESO http://www.eso.org/public/images/archive/to100/)

measurements of our galactic center with a high spatial resolution [3]. In these extensive measurements which took several years, the positions of stars and their velocities have been determined very accurately (Fig. 7.21). These measurements have been performed with infrared light at $\lambda = 2.5\ \mu$m . In order to achieve the demanded spatial resolution, the very large telescope of the ESO at the Paranal has been used, which is equipped with active and passive optics (see Sect. 2.x) and reaches an angular resolution of better than $0.02''$.

For stars on a circular orbit with radius r and the velocity v(r) the gravitational force equals the centripetal force. This gives the equation

$$m \cdot v^2/r = G \cdot M_r \cdot m/r^2 \tag{7.22}$$

where M(r) is the total mass inside the radius r. The resultant mass is

$$M_r = v^2 \cdot r/G. \tag{7.23}$$

The experimental results (Fig. 7.22) show that the enclosed mass, obtained from Kepler's law stays constant up to the radius $r = 0.01$ pc from the galactic center. It has a numerical value $M = 2.41 \cdot 10^6\ M_\odot$. The closest star which could be still resolved had a distance $r = 0.01$ pc from the center. This means that the central mass is concentrated in a volume with radius $r \leq 0.01$ pc. The mass density must be therefore

$$\begin{aligned} \varrho_m &> 2.61 \cdot 10^6 M_\odot / \left(4\pi \cdot 10^{-6}\ \mathrm{pc}^3/3\right) \\ &= 1.7 \cdot 10^{-6}\ \mathrm{kg/m}^3. \end{aligned} \tag{7.24}$$

According to the Jeans- Criterion a mass distribution with such a high mass density is gravitational unstable and collapses. The mass M is far above the stability limit of a

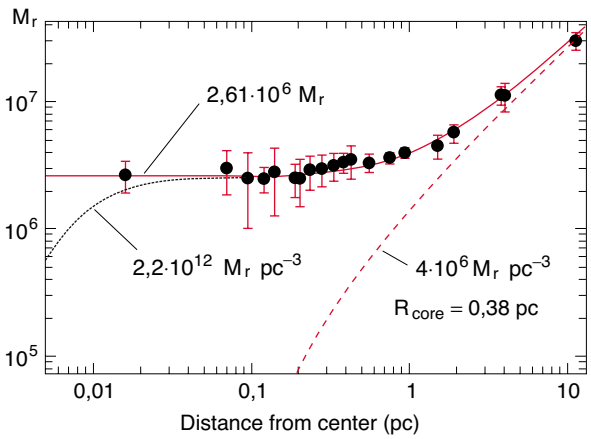


Fig. 7.22 Central mass inside the radius r as a function of the distance r from the center. The points give the experimental values of the central mass obtained from the Kepler motions of stars close to the black hole

neutron star. This implies that the collapse must lead to a black hole. The Schwarzschild radius is

$$\mathcal{R}_S = 7.7 \cdot 10^9 \mathrm{m}$$

(see Sect. 2.y).

Further hints to the existence of a super-massive black hole are obtained from measurements of angular resolved γ-radiation from the center of our galaxy, which have been performed by the HESS-collaboration with the HESS Telescope in Namibia. HESS is the acronym for "high

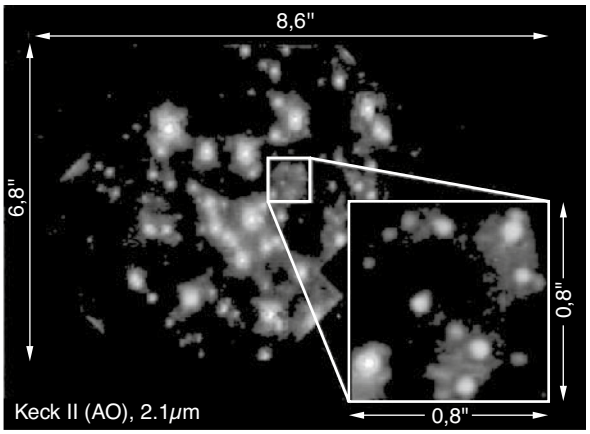


Fig. 7.23 Compact star cluster around Sagittarius A* [6]. The individual stars are resolved due to the adaptive optics on the large Keck-telescope. The motion of the stars becomes visible on time scales of several months (Keck Observatory, Mauna Kea, Hawaii)

energy stereoscopic system". This system consists of 4 spatially distributed Cerenkov detectors, which detect the energy and the direction of γ-radiation with γ-quantum energy above 40 GeV. The measured narrow angular distribution and the energy spectrum indicates a locally narrow acceleration mechanism, which might be a black hole that accelerates matter into its center and brings it to very high energies, emitting gamma radiation.

These measurements are the first experimental proof, that really a black hole sits in the center of our galaxy. Stars and dust with radius $r > r_s$ around this black hole with radius r_s can be brought by collisions into regions with $r < r_s$ and are then swallowed by the black hole. First hints that such a mass mug really takes place are recordings of radiating clouds in the proximate surroundings of the galactic center, which have been taken with the Keck-telescope (Fig. 7.23). One can recognize a gas cloud which is pulled on spiral orbits towards the galactic center.

7.10 The Dynamics of Our Galaxy

Although our galaxy has been formed already about 10^{10} years ago, it still is changing continuously. Not only because always new stars are born and old stars die (they become white dwarfs, neutron stars or black holes), or why the massive black hole in the center "eats" stars in its surroundings, but also because our galaxy interacts with several other small galaxies in our neighbourhood and attracts stars and gas from them. The continuous influx of gas could be recently proved by the discovery of hydrogen gas clouds which move with high velocities through the outer regions of our galaxy. Their extension can reach up to 10.000 light years and they contain a total mass of up to $10^7 M_\odot$. Measurements of Doppler-shifts of these fast clouds show that clouds which move away from the glaxtic center are observed as well as those which are attracted towards the center.

It seems that our galaxy collects huge amounts of gas from outside and also ejects about the same quantity of gas. Further discoveries show that our galaxy is surrounded by a sphere containing a hot thin plasm which extends far beyond the region assumed so far. Also the halo around our galaxy contains such a plasm.

There are different explanations of these fast gas clouds. Some astronomers assume that these clouds are from small galaxies circulating around our galaxy, which, because of its much larger mass attracts the gas and steals it from the dwarf galaxies during their closest approach. In former times much more dwarf galaxies had surrounded our galaxy, but in the course of time it has annexed many of them and therefore has increased its mass. There are indications that these gas clouds consist of a mixture of H and H_2 and of dark matter.

Similar conditions are certainly also present for other galaxies because our galaxy is no exception.

These considerations show that not only stars but also whole galaxies are subject to continuous changes which corrects the old perception of an eternal static non changing universe.

7.11 The Space Between the Stars; Interstellar Matter

The interstellar space between the stars is not empty but contains gas (atoms, molecules, ions), dust particles and electric and magnetic fields, and of course also photons emitted by the stars. From the measured polarization of radio radiation emitted from far away radio sources, which is transmitted through our galaxy bevor it reaches the earth, and also from the polarization of light emitted by stars of our galaxy, one can imply the strength and the direction of the galactic magnetic field. It turns out that this field varies strongly for the different locations. It has for instance in the spiral arms a different strength and direction than in the central galactic disc. Its mean strength B is about $5 \cdot 10^{-6}$ Gauß ($5 \cdot 10^{-10}$ Tesla). This is very small compared with the mean magnetic field of the earth ($B = 20\ \mu T = 2 \cdot 10^{-5}$ T).

The charged particles of the cosmic radiation, which penetrate into the galaxy are forced by the magnetic field onto circular paths, where they emit synchrotron radiation, with a spectrum ranging from radio frequencies up to the x-ray region.

The interstellar matter in our galaxy [3] is, like in other spiral galaxies, not uniformly distributed. Besides regions with low density there are matter concentrations in gas- and dust- clouds which are the birth regions of new stars. The main fraction of the total mass of the interstellar matter is

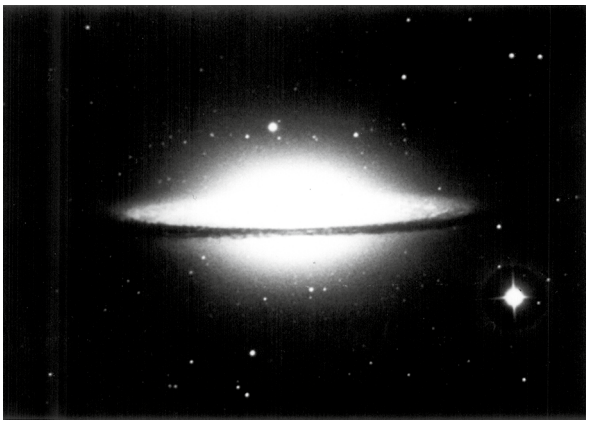


Fig. 7.24 Edge-on View of the galactic disc of the spiral galaxy M104 (NGC4594) in the Virgo-galaxy cluster. The absorbing dust in the disc appears as dark central band (Wikipedia)

concentrated in a flat disc in the galactic plane. The thickness of this disc is smaller (50–100 pc) than that of the star distribution, because the stars have vertical velocities and therefore spend part of their time at larger z-values from the galactic plane z = 0. This interstellar matter in absorbing clouds can be seen in other galaxies as a dark thin band if they are oriented in such a way that we look at them side on (Fig. 7.24). The cloud appears dark, because it absorbs the light of stars behind the cloud.

In our galaxy the mass of the total interstellar matter amounts to about $10^{10}\ M_{\odot}$. This is about 10% of the total mass of the galaxy. It consists to 99% of gas and only about 1% of dust. The density of the gas component is about 1 atom per cm^3, that of the dust component about 1 dust particle per $10^{12} cm^3$. Meanwhile it has been found that the total halo of our galaxy is filled with gas of small density [4].

The dust component consists of small micro-particles with diameters $D \approx 0.01-1\ \mu m$, which is within the order of magnitude of the wavelength λ of visible to ultraviolet light. Therefore Mie-scattering (see Vol. 2, Sect. 8.3) of the star radiation is the pronounced scattering mechanism. Since the scattering cross section is wavelength dependent ($\sim 1/\lambda^4$) the star light passing through such clouds is shifted towards longer wavelengths. A Star behind such an absorbing cloud appears not only weaker but also more reddish. The mean density of dust particles in the interstellar space is only 10^{-6} particles per m^3, it can be much higher in dust clouds.

Although the fraction of the gas component in the interstellar matter is 99%, it contributes less to the absorption than the 1% of the dust component. The reason for this is that the gas (mainly H-atoms and a small fraction of CO, HO, CN, HCN molecules) absorbs only at discrete wavelengths, corresponding to the transitions between two atomic or molecular levels, while the dust component absorbs and scatters light of all wavelengths.

The main part of information about interstellar matter stems from radio-astronomy. In particular the measured intensity of the 21 cm line (corresponding to the hyperfine transition $F = 0 \rightarrow F = 1$ in the $1^2S_{1/2}$-ground state of the H-atom) gives information about the density of hydrogen atoms in our galaxy. In dense cold hydrogen clouds this line can be noticeably absorbed, in spite of the small absorption coefficient. Measurements of the Doppler-shift of this line gives information about the velocity of the absorbing gas clouds Fig. 7.10).

The 21 cm line can be observed either in absorption (if the F = 0 level is more populated than the F = 1 level), or in emission, if the upper hfs-component F = 1 is populated by collisions or by fluorescence from the excited H*(2p) atoms. The F = 1 component can be also populated by photo-dissociation of H_2-molecules. Therefore the 21 cm line can be observed in emission, if the proper conditions are present in the interstellar matter.

One distinguishes between a cold component of the interstellar matter (T = 10–20 K) which is mainly concentrated in cold molecular clouds, and a hot component (T = 5.000–10.000 K), which has only a small density and consists mainly of atoms and ions (Table 7.5).

The hot component is heated by absorption of star radiation and of cosmic radiation. The ultraviolet radiation from stars can ionize the neutral gas component and can also release electrons from dust particles.

Cold gas clouds can partly absorb continuous radiation from hot stars behind them [4]. In such cases one observes an absorption line in a continuous background emission spectrum.

In hot clouds, which are heated by the UV-radiation from stars, excitation and ionization of neutral atoms and dissociation of molecules occur. Such hot clouds, which are named H II regions after their main component $H^+ + e^- (H^+ = H\ II)$, emit light and therefore appear as impressive bright reddish to blue extended regions (up to 100 pc) The light emission is mainly due to the recombination radiation generated in the process $H^+ + e^- \rightarrow H^*$, where excited atomic states H^* are populated. These excited

Table 7.5 The components of the interstellar medium

Component	Temperature/K	Density (particles/cm^3)	Volume %
Molecular clouds	10–20	10^2–10^6	<1
Cold neutral medium	50–100	20–50	1–5
Hot neutral medium	5000–10.000	0.2–0.5	10–20
Hot ionized medium	7000–10.000	0.2–0.5	20–50
Very hot ionized medium	10^6–10^7	10^{-4}–10^{-2}	30–70

atomic states can again emit line radiation when decaying into lower states. For excited hydrogen atoms the Balmer α-radiation ($n = 3 \rightarrow n = 2$) has by far the highest intensity of all visible radiation. It gives the Orion Nebula (Fig. 5.18) or the Trifid nebula (Fig. 7.25) their reddish appearance.

The molecules in the interstellar medium have been discovered mainly by radio astronomy. The Doppler-shift of the molecular lines gives a very convenient means to measure the velocity of interstellar matter and its spatial variation. For example, colliding gas clouds in the Orion nebula are at present intensively investigated with this molecular radio astronomy [5].

Rotational transitions of polar molecules (e.g. CO; HO; H_2O) results in microwave radiation, which is emitted or absorbed, depending on the population difference $\Delta N = N_u - N_l$ between upper and lower level of the transition. For $\Delta N > 0$ emission occurs, for $\Delta N < 0$ absorption.

The absorption coefficient $\alpha(\nu)$ is related to the absorption cross section $\sigma(\nu)$ for the transition $E_k \rightarrow E_i$.

with $E_k < E_i$ as

$$\alpha(\nu) = \left[N_k - (g_k/g_i)N_i\right]\sigma(\nu), \tag{7.25}$$

where the factors g_i, g_k are the statistical weights of the levels E_i and E_k. Absorption occurs for $\alpha(\nu) > 0$, i.e. for $N_k > (g_k/g_i)N_i$.

Under thermal equilibrium at the temperature T the population distribution in the different levels is governed by the Boltzmann distribution

$$N_i/N_k = (g_i/g_k) \cdot e^{-\Delta E/kT} \tag{7.26}$$

and the absorption coefficient

$$\alpha(\nu) = N_k\left[1 - e^{-\Delta E/kT}\right] \cdot \sigma(\nu) > 0 \tag{7.27}$$

is always positive, which means that absorption occurs (Fig. 7x).

Since higher excited states can be populated by absorption of star light, these states can emit fluorescence on transitions into lower states. It turns out, that under certain conditions the fluorescence populates preferentially the upper microwave state which therefore gets a larger population than the lower one. In such cases the absorption coefficient becomes negative and amplification of the microwave radiation occurs instead of absorption. The star light in combination with collisional processes acts as a pump to produce population inversion. One observes a large amplification of the microwave radiation. This phenomenon is called an **astronomical microwave maser**.

Such cosmic microwave masers produce a strongly enhanced microwave radiation on rotational transitions in molecules or on fine-structure transitions. For example, in 1963 extremely strong microwave radiation was found at $\lambda = 6\,\text{cm}$ and $\lambda = 18\,\text{cm}$ emitted by gas clouds close to the galactic center. It could be assigned to rotational transitions between

Fig. 7.25 The Trifid nebula in the constellation Sagittarius Archer, M20 or NGC6514) is the superposition of an open star cluster, a trifurcated absorption nebula and an emission nebula, which is excited by the radiation from a white giant star. Its diameter is 5 pc, its distance 1000 pc (*Photo* Richey-Cheretion Reflector, Haman Observatory Salt Lake City)

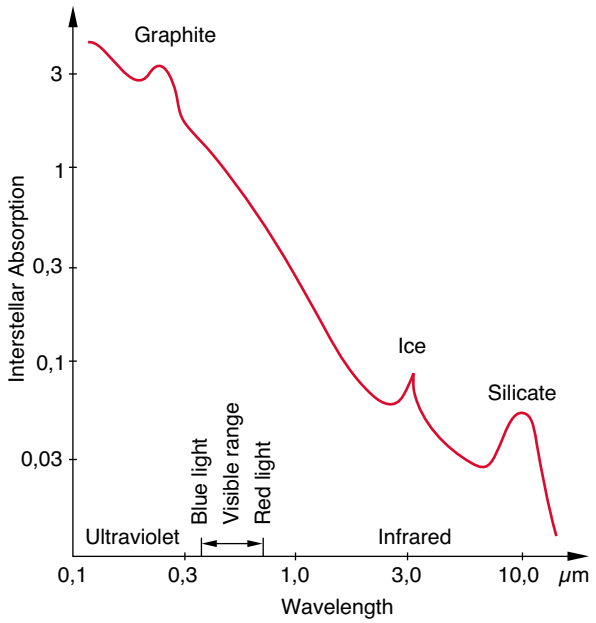


Fig. 7.26 Spectral dependence of the absorption coefficient α(λ) of the interstellar dust in the spectral range 0.1–10 μm (Lexikon der Astronomie, Spektrum Akademischer Verlag Heidelberg Germany)

Fig. 7.27 Dark molecular clouds in the Eagle nebula M16 (ESA Hubble CCBY 3.0. https://commons.wikimedia.0rg/w/index,php?curid=37,706,197)

different rotational levels of the OH-radical and was called the OH-maser. Andresen et al. [5] showed by laboratory experiments, that the selective population of one Λ-component of the rotational levels in the $^2\Pi$ ground-state of OH occurs at the photodissociation of H_2O-molecules by UV-light. This produces an inversion and is responsible for the maser action.

The question, how molecules are formed in these gas clouds has been long controversely discussed. The collision between 3 atoms to form a diatomic molecule (the third atom is necessary to take away the relative kinetic energy of the two colliding atoms to form a stable diatomic molecule), is extremely unprobable because of the low density (n = 1 atom/cm^3). Therefore, it is nowadays assumed that the formation of molecules occurs on the surface of dust particles, where two atoms can recombine. The stabilization of the molecules is here possible because the excess energy can be transferred to the dust particle. Although the density of dust particles is very low ($n = 5 \cdot 10^{-6}/\mathrm{m}^3$), they play the essential role for the formation of molecules. Even larger molecules which contain up to 12 atoms, were found.

The importance of the interstellar dust for the development of stars and galaxies was for a long time underestimated. Only recently detailed investigations have shown, that dust clouds inside a gas nebula can shield the radiation within the gas nebula, thus decreasing the radiation pressure and facilitate the collapse of a gas cloud.

In cold dense gas clouds ($n > 10^3$ atoms/cm^3) the direct formation of molecules by collisions of 3 atoms becomes more probable. These clouds are the birth places of new stars (Fig. 7.27).

In Table 7.6 a selection of molecules is listed, found up to now in gas clouds.

In the last years also infrared measurements have been started which give complementary information about the interstellar matter. In particular the absorption and scattering of infrared and visible light by dust particles is a very sensitive method for the determination of the size and the composition of dust particles.

In Fig. 7.26 the absorption curve $\alpha(\nu)$ of typical dust particles in cold dust clouds is plotted in the spectral region $\lambda = 0.1-10\,\mu\mathrm{m}$.

Meteorites and Comets contain small amounts of interstellar dust. Therefore, the investigation of the atomic composition of the dust particles with sensitive mass spectrometers give information about the star which had expelled the dust into its surroundings. The main part of the interstellar dust comes from red giants, which blow up at the end of their life and eject part of their envelope (see Sect. 5.x). For example, carbonaceous meteorites have gathered their carbon from stars which had been stable until the carbon-burning started and then had been blown up to red giants. The carbon had been transported by convection from the core to the envelope and was then ejected into space.

Table 7.6 Selection of molecules discovered up to now in the interstellar gas- and dust clouds (Lexikon der Astronomie Springer Heidelberg 1995)

2 Atoms	3 Atoms	4 Atoms	5 Atoms	6 Atoms	7 Atoms	8 Atoms	9 Atoms	11 Atoms
H_2	H_2O	NH_3	H_2CNH	CH_3OH	CH_3NH_2	$HCOOCH_3$	CH_3OCH_3	HC_9N
CH	H_2S	C_2H_2	CH_2CO	CH_3CN	CH_3C_2H	CH_3C_3N	C_2H_5OH	
CH^+	C_2H	H_2CO	H_2NCN	$HCONH_2$	CH_3CHO		C_2H_5CN	
C_2	HCN	H_2CS	HCOOH	CH_3SH	C_2H_3CN		HC_7N	
CN	HNC	HNCO	C_3H_2		HC_5N		CH_3C_4H	
CO	HCO	HCNO	C_4H					
CO^+	HCO^+	C_3N	HC_3N					
CS	HOC^+	C_3H	SiH_4					
CS^+	HN_2^+	$HOCO^+$						
NS	OCS	C_3O						
OH	SO_2	HCNS						
SO	SiC_2							
SiO	HNO(?)							
SiS	HOC^+(?)							
NO								
NaCl								

Summary

Our own galaxy (milky way) contains about 10^{11} stars. It is a spiral galaxy with two spiral arms and a bar and can be classified as Sb bc, consists of a flat disc (diameter 30 kpc, thickness 0.6 kpc), with a bulge around its center, 2 spiral arms, where new stars are born, and many globular clusters in the halo around the center of the galaxy.

The galaxy rotates differentially, i.e. the angular velocity ω(R) depends on the distance R from the center.

The revolution period is about 200 million years.

The spiral arms are probably formed by density waves caused by gravitational perturbations of the cylindrically symmetric mass distribution.

Between the stars the interstellar medium exists, consisting of gas and dust. Its total mass is about 10% of the total star mass in the galaxy. The gas component amounts to about 99%, the dust component only to 1% of the interstellar medium. One distinguishes between a cold interstellar medium (T ≈ 10–20 K) and a hot component (T ≈ 5.000–10.000 K).

In the center of our galaxy a black hole has been found with a mass of 2.4 million $M_\odot$. Its Schwarzschild radius is $R_S = 1.2 \ldots 10^{10}$ m = 2% of the sun radius.

If by specific collision processes a population inversion between the hyperfine levels of atoms or between rotational levels of molecules in the interstellar medium is generated, cosmic maser action can occur which leads to a strong microwave amplification on specific wavelengths.

In the disc of our galaxy many open star clusters exist, consisting of 10–1000 stars which have all been formed at the same time by the collapse of a large gas-and dust cloud.

Problems

7.1 How large is for a homogeneous density n(M) of stars with the absolute magnitude M the ratio N_{m+1}/N_m of observed galaxies with apparent magnitude m + 1 resp. m?

7.2 Derive Eq. (7.18).

7.3 What is the spectral width of a radio signal, observed along a radial line of sight from the earth to the center of our galaxy, if the density ϱ of the emitting molecules increases as $\varrho = \varrho_0$(1-r/R) with decreasing distance r from the center?

7.4 The globular cluster M13 in the constellation Hercules appears under the angle $\alpha = 25$ and has the apparent magnitude m = 5.8. The variable stars in M13 have the absolute magnitude M = 0.0 and the apparent magnitude m = 14.16. What is the distance of M13 and how large is its diameter?

References

1. R. Kurth; Introduction to Stellar Statistics (Elsevier Amsterdam 1968)
2. https://solarsystem.nasa.gov/missions/gaia/in-depth
3. https://en.wikipedia.org/wiki/Globular_cluster
4. Supermassive Black Hole Sagittarius A* | NASA
5. The Story of How Scientists Discovered the Black Hole at the Milky Way's Centre - The Wire Science Physics Nobel honours: Discoveries about black holes | Science | AAAS
6. H. Scheffler, H, Elsässer: Physics of the Galaxy and Interstellar Matter (Springer, Berlin, Heidelberg 1987) E. Krügel: An Introduction to the Physics of Interstellar Dust (Taylor and Francis, New York 2008)

7. NASA - The Milky Way's Hot Gas Halo
8. I. Gissis, A. Fisher and E, Behar: X-ray absorption by Cold Gas. Simulating interstellar Molecular Clouds in the Lab. (Phys. Rev. **E104**, 015205 (2021) https://www.semanticscholar.org/paper/H2%2C-CO%2C-and-Dust-Absorption-through-Cold-Molecular-Lacy-Sned
9. Huge Cloud Of High Temperature Gas Found In Orion Nebula—ScienceDaily
10. Cosmic masers - NASA/ADS (harvard.edu)

The Dark Universe

8

Contents

8.1 Dark Matter 233
8.2 Dark Energy 235
Summary 236
Problems 236
References 237

The important question of the real mass density in the universe, which is essential for the determination of the age of the universe, and its future, has got new intellectual nourishment after new precise measurements of gravitational forces in galaxies and galaxy clusters have been performed. The results of such measurements showed, that the mass obtained from the radiation emitted by stars or galaxies must be only a small percentage of the total mass. In order to explain all measurements up to now, one has to introduce additional mass and energy parts, which do not appear in the emitted radiation. Since they cannot be seen they are called "*dark matter*" and "*dark energy*".

8.1 Dark Matter

The first indication of the existence of dark matter was assumed by Lord Kelvin, who estimated already 1884 the number of dark bodies in the Milky Way, necessary to explain the velocity distribution of stars in our galaxy, which could not be explained only by the visible stars.

Later on the first indications of the existence of dark matter came from measurements of the angular rotational velocity of stars in a galaxy as a function of the distance from the galactic center (see Sect. 6.4), Calculating the rotational curves one obtains according to the Keppler laws a much larger mass of the galaxy, than determined form its luminosity. There must be therefore a mass which does not contribute to the emitted radiation but acts via gravitational forces on the visible matter. The assumption of a black hole at the center of the galaxy can also not explain the measured rotation curves. The analysis of typical rotation curves $\omega(r)$ (Fig. 8.1) reveals that the dark matter must be distributed uniformly within and outside the visible matter. The visible matter of the galaxy must be embedded into a halo of dark matter. This dark matter does not absorb or emit electromagnetic radiation, but it does interact by gravittional forces with visible matter. What are the basic massive particles of dark matter is up to now one of the large open questions of astrophysics.

Another indication of the existence of dark matter is the stability of galaxies in a cluster and the velocity distribution of galaxies in such clusters. Taking into account only the visible matter would result in a limited stability, which could not keep the different galaxies in the cluster together for a longer time.

A very convincing argument for dark matter is the early formation of stars and galaxies during the development of the universe. Since the radiation pressure had prevented mass accumulation in the era of the radiation dominated universe ($t<370.000$ years) it should have taken much longer to form massive objects such as stars and galaxies. On the other side the dark matter was not hindered to cluster by the radiation, because it does not interact with it. Therefore it could cluster already at an earlier time and attract the visible matter to form density maxima which acted as the birth places for stars. From the fluctuations of the cosmic background radiation (Fig. 8.2) one can infer presicely the time of the formation of mass density maxima. The dark matter must have existed already at the beginning of the universe before the inflationary phase. Quantum fluctuations must have been the seeding structures for later mass density fluctations, because during the inflationary phase these quantum fluctuations had been blown up to macroscopic density fluctuations.

There are quite a lot of assumptions about the possible basic constituents of dark matter.

Initialy it was assumed that neutrinos might be the basic particles of dark matter. More accurate measurements of upper limits for the neutrino mass ($m(\nu_e)+m(\nu_\mu)+m(\nu_\tau)<0.4\ \mathrm{eV/c^2}$) however excluded this assumption [36].

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_8

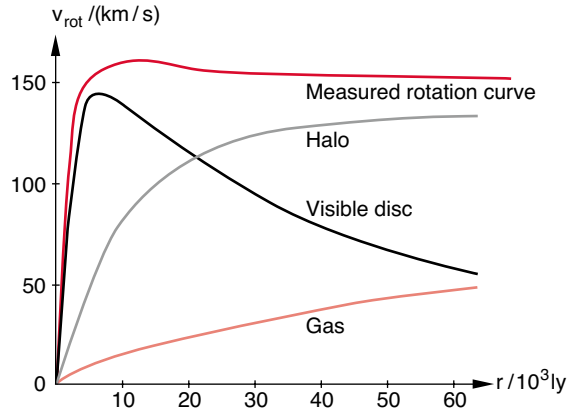


Fig. 8.1 Measurement of a typical rotation curve (red) and the contributions of the visible disc (black curve) and the (Halo) including the interstellar gas

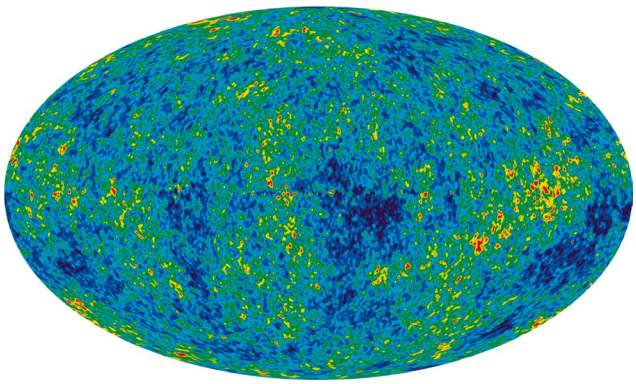


Fig. 8.2 A map of the tiny fluctuations of the cosmic background radiation [By NASA/WMAP Science Team. http://map.gsfc.nasa.gov/media/121238/ilc_9yr_moll4096.png, Public Domain]

A further hypothesis assumes *MACHOS* (massive compact halo objects) as building block of dark matter. These should be massive, not detailled specified objects in a galaxy, which are distributed inside and around a galaxy like a halo. If this are extended solid objects, they should consist of elements beyond hydrogen and helium. However, such elements have been formed only later on by fusion processes in stars and cannot have been present before the formation of stars.

Although most galaxies host a massive black hole in their center this cannot explain the dark matter because of two reasons: The measured rotation curves of galaxies cannot be reproduced by the assumption of a black hole. Furthermore the mass of even a massive black hole represents less than 1% of the mass of the whole galaxy, while the mass of the dark matter must be larger than the visible mass.

A further presumption assumes brown dwarfs (final states of small stars where no nuclear fusion processes had taken place) and cold solidified white dwarfs as constituents of dark matter. Estimations of their number let their relevance for dark matter however appear quite unlikely. Moreover, they could not explain the rotational curves and the stability of galaxy clusters.

Possible candidates are **WIMPS** (weakly interacting masive particles). These are assumed massive subatomic particles with masses much larger than the proton mass, but only very weakly interacting with particles of ordinary matter. At the large particle accelerators an intense search for such particles is underway. This search is difficult, because of the weak interaction (only by gravitational or weak forces), which makes the scattering cross sections for collisions with ordinary matter extremely small. The idea of their detection is based on an energy balance: If two protons with a known energy collide anticollinear, they produce new particles. The barionic and mesonic particles and their energy can be measured. If there is any missing energy it must be due to non-interacting particles wich are not caught by the detectors. Their total energy (rest energy + kinetic energy) equals the missing energy and might be attributed to WIMPs.

Another detection technique relies on large germanium or silicon crystals, cooled down to temperatures $T < 1$ K, which produce either a light pulse or an energy signal if a particle passes through the crystal. Even if the interaction of WIMPs with the crytal atoms is very weak, the probability to detect them is not zero. Estimations show that about 1 event per year and kg detector material might be expected. The detectors must be therefore very large. This is similar to the detection of neutrinos, which also have a very weak interaction with matter, but still could be detected with a sophisticated equipment.

All astronomical observations provide, however, convincing evidence that this dark matter really exists, otherwise the standard model has to be revised. Possible appearance forms of dark matter are cold and hot dark matter. Cold dark matter had such small velocities in the early universe that it could not reach distances of a proto-galaxy, i.e. its streaming length iss maller than the diameter of a proto-galaxy. Most experimental observations can be best explained by the model of cold dark matter.

The Israelic cosmologist *M Milgrom* from the Weizmann Institute published a hypothesis, that the gravitational force might not be strictly proportional to $1/r^2$. The gravitational acceleration might be f or ranges below a value 10^{-6} m/s^2 only proportional to $1/r$ instead of $1/r^2$. With this assumption the rotational curves of many galaxies can be explained quite well, without the assumption of dark matter. However, many other observed phenomena, such as the early structure formation or the velocitiy distribution of galaxies in clusters can not be explained by this hypothesis. Therefore it is no longer regarded as a true explanation.

The dark matter constitutes about 85% of all matter in the universe. This means that the visible matter represents only 15% and is therefore by far the minor fraction of the total matter.

From these considerations one can see, that there are still many open questions, For satisfactory answers many more experiments and observations are necessary and perhaps also a revision of our momentary model of the universe should be considered.

8.2 Dark Energy

Recent accurate distance measurements of supernovae Ia covering a large range of distances, have shown the surprising result, that the escape velocity of the galaxies increases with time. This means that the universe expands accelerated, contrary to our former belief, that the expansion should slow down with time, because the gravitational attraction between the galaxies should act as a retarding force to slow down the expansion velocity. If we do not want to give up well tested physical laws (for instance Newtons gravitational law) we must introduce a new repulsive force which over-compensates the attractive gravitational force and can explain the increasing escape speed. The energy, related to this repulsice force is called "*dark energy*". It is proportional to the cosmological constant Λ in Sect. 6x). Its mass equivalent of the energy density is estimated as $1 \cdot 10^{-27}$ kg/m^3. This is smaller than the density $\rho_c = 9 \cdot 10^{-27}$ kg/m^3 of the ordinary matter. However, the dark energy fills every part of the universe, even the empty space where no ordinary mass is present. Therefore the total dark energy is larger than the energy equivalent of the ordinary matter. The introduction of the dark energy implies that we need a new concept of the development of our universe which includes the dark energy.

Up to now nobody knows what the dark energy really is and how to include it into a new concept of the universe. There had been theoretical attempts to explain the dark energy as vacuum energy. But the numerical results of such calculations deviate by many orders of magnitude from the experimental value, that would fit the measured accelerated expansion. There are several other attempts to explain the dark energy, which are, however, controveresely discussed. This problem therefore represents an interesting new field of astrophysics.

According to theoretical considerations the dark energy density ρ_{dE}, should be constant, i.e. the total dark energy $E_{dE} = \int \rho_{dE} dV$ should increase in the expanding universe with the scale parameter R proportional to R^3. During the first billion years the mass densitiy was still larger than the dark energy density and the expansions was slowed down. After a critical time t_0 the dark energy density predominated and the expansion started to accelerate (Fig. 6.24).

More detailled angular resolved measurements of the cosmic background radiation by the Satellites KOBE (cosmic background explorer) (Fig. 8.3) and WAMP (Wilkinson Microwave Anisotropy Probe) showed a tiny anisotropy, which is related to spatial structures of mass and energy density in the universe at the time when matter and radiation were decoupled (about 370.000 years after the big bang). These very small variations can be.

represented by a multipole expansion of the intensity $I(\alpha, \delta)$ which depends on the galactic coordinates α and δ. This is similar to the multipole expansion of the electric field generated by an extended charge distribution, into multipoles of the order l (l = 1 = monopole, l = 2 = dipole, l = 2 Quadrupole etc. They give information about the spatial oscillations of the mass density at the time of the decoupling of matter and radiation [20]. During the contraction of the matter the radiation (photon gas) had a higher temperature, during the expansion it was colder. The maximum of the temperature deviation from the mean value corresponds in the multipole expansion of the background radiation to the term for l = 200. As is shown in [20] the intensities $I(l)$ of the different terms in the mutipole expansion allow the estimation of the percentage of the barionic matter as $\Omega_B = \rho_B/\rho_t = 0.04$ where ρ_t is the total mass density (including the dark energy divided by c^2). The fraction of the dark matter is, according to this estimation, $\Omega_{DM} = \rho_{DM}/\rho_t = 0.21$. From the measured acceleration of the expansion one can estimate the fraction of the dark energy as $\Omega_{DE} = 0.74$ (Fig. 6.16 and 8.4). The total sum is $\Omega = \Omega_B + \Omega_{DM} + \Omega_{DE} = 0.99$.

The missing 1% is due to the mass of the neutrinos and the mass equivalent $m_{BR} = E_{BR}/c^2$ of the background radiation. In Fig. 8.4b the weighted deviations Δ of the expansion parameter from the Hubble law are plotted for different percentages of the dark energy. The red points represent the experimental observations. The black curve for $\Omega = 0$

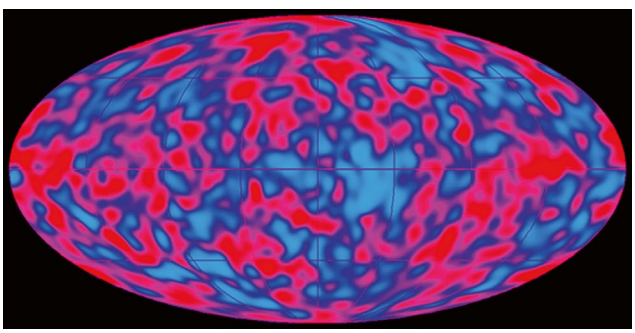


Fig. 8.3 Temperature fluctuations of the cosmic background radiation measured by the satellite COBE The galactic center is in the middle, the galactic plane a horizontal line through the center of the picture, Orion is to the right and cygnus to the left The map is a resupt of 2 years observation. (NASA)

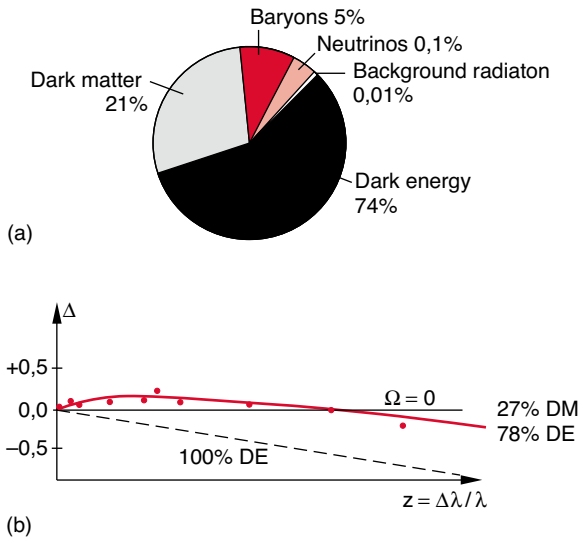


Fig. 8.4 **a** The different shares of baryonic matter, dark matter and dark energy to the total energy. **b** Deviations of the expansion velocity from the Hubble law for different contributions of dark matter and dark energy. The red points represent the experimental values from many distance measurements based on observations of supernovae Ia at different distances

corresponds to an empty universe, the dahed curve shows the situation for a universe with 100% dark energy. The best fit to the measured points is the curve for $\Omega_{DE} = 0.74$ i.e. with 74% dark energy.

There is only indirect evidence for the existence of dark energy. It is based on three independent sources:

(1) The critical mass density to form an observable flat universe cannot be reached only by ordinary and dark matter. There must be an additional contribution which might be given by the mass equivalent of the dark energy. Measurements of the slight anisotopy in the microwave background radiation show, that the universe must be close to a flat universe.
(2) The observed accelerated expansion demands a repulsive force, which cannot be provided by ordinary and dark matter, because both result in attractive forces. There must be a strong repulsive force, which is attributed to the dark energy. The acceleration of the expansion can be also determined from observations of the voids (empty space) between the galaxy clusters. Their dependence on the age of the surrounding galaxy clusters leads to the conclusion of an accelerated expansion of the universe.
(3) Measurements show that in the universe there exist large scale wave patterns of the mass density. They can be explained by assuming local variations of the dark energy.

More information can be found in [1–8].

Summary

- The rotational curves of galaxies indicate that there must be more matter than deduced from the baryonic visible matter. This additional matter is called *dark matter*.
- The observed velocity distribution of galaxies in galaxy clusters can be only explained, if dark matter is accepted.
- The early formation of structures in the universe can not be explained, if only visible matter would be present.
- Before the decoupling of radiation and matter baryonic matter could not cluster because the radiation pressure prevented it. Dark matter is not affected bye el. magn, radiation and could therefore cluster already much earlier.
- According to our present understanding the clustering started already as quantum fluctuations of the energy density in the very early universe. These quantum fluctuatuions were blown up to macroscopic density fluctuations during the inflationary phase of the universe.
- The dark energy has been postulated after the discovery of the accelerated expansion of the universe, which was measured by the distances of supernovae Ia, which served as accurate distance candles. The dark energy acts as repulsive force.
- The energy density of the dark energy is assumed to be constant during the expansion. This means that its total amount increases with the volume of the expanding universe.
- At present nobody knows what the basic foundations of dark matter and dark energy really are. Intense investigations have not yet found any particle which could be responsible for the dark matter, nor any accepted model of the dark energy has been developed.

Problems

8.1 Assume a uniform mass density in a galaxy.
(a) What would be the rotation curve $\omega(r)$, according to Kepler's laws?
(b) How would $\omega(r)$ change for a density function $\rho(r) = \rho_0(1 - r/R)$ where R is the radius of the galaxy disc with a mass distribution $m(r) = m_0(1 - r/R)$.
(c) What would be the rotation curve, if the whole mass of the galaxy would be compressed into a point at $r = 0$?

8.2 Assume that the energy density of the dark energy is $\rho_{dE} = 7 \times 10{-}^{27}$ kg/m^3 uniformly distributed within the universe. What would be the acceleration of a galaxy with and without the dark energy when the mass density of the baryonic mater is $\rho = 6H_0^2/(8\pi G)$, twice the critical density.

References

1. I. Nicolson: Dark Side of the Universe. (Canopus Publ. Bristol 2007)
2. https://de.wikipedia.org/wiki/Dark_Matter
3. https://www.nasa.gov/audience/forstudents/9-12/features/what-is-dark-matter.html
4. Sources and Detection of Dark Matter and Dark Energy in the Universe. Fourth International Symposium Held at Marina del Rey, CA, USA February 23–25, 2000. **Cline**, David B. (Ed.)
5. The Invisible Universe: Dark Matter and Dark Energy Ed. By Eleftherios Papantonopoulos: Lecture Notes in Physics 720 (Springer, Heidelberg 2007)
6. https://en.wikipedia.org/wiki/Dark_energy
7. Riess, Adam G.; Filippenko; Challis; Clocchiatti; Diercks; Garnavich; Gilliland; Hogan; Jha; Kirshner; Leibundgut; Phillips; Reiss; Schmidt; Schommer; Smith; Spyromilio; Stubbs; Suntzeff; Tonry (1998). "Observational evidence from supernovae for an accelerating universe and a cosmological constant". Astronomical Journal. 116 (3): 1009–1038. arXiv:astro-ph/9805201. Bibcode:1998AJ....116.1009R. https://doi.org/10.1086/300499. S2CID 15640044.
8. "Measurements of Omega and Lambda from 42 high redshift supernovae". Astrophysical Journal. 517 (2): 565–586. arXiv:astro-ph/9812133. Bibcode:1999ApJ...517..565P. https://doi.org/10.1086/307221. S2CID 1

The Evolution of Our Solar System

9

Contents

9.1 The Origin of the Elements 239
9.2 The Origin and Development of Our Solar System 241
9.3 Separation of Gases and Solid Matter. 244
9.4 The Age of Our Solar System 245
9.5 Other Solar Systems and Exoplanets 248
9.6 The Development of the Earth 250
Summary 257
Problems 257
References 257

The existence of heavy elements in our solar system prooves that it cannot have been formed already in the early phase of the universe, because these heavy elements had been only produced later on by fusion processes in stars. Therefore our solar system must be at least of 2nd or probably of a later generation in the development of stars and planets in our universe.

9.1 The Origin of the Elements

In Sects. 9.6 and 9.3 we have discussed that during the early phase of the univerese i.e. in the first 4 min after the big bang only the light nuclei $^{1}H^{+}$; $^{2}{}_{1}H^{+}$; $^{4}{}_{2}He^{++}$ and Li^{+++} had been synthesized (Fig. 9.1).

Nuclei, which are heavier than $^{4}{}_{2}$He could barely have been produced, because there are no stable nuclei with 5 or 8 nucleons. For instance at collisions of $^{4}{}_{2}$He with protons p or neutrons n or $^{4}{}_{2}$He nuclei the following elements are generated:

$$\begin{aligned} {}^{4}_{2}\mathrm{He} + \mathrm{p} &\rightarrow \\ {}^{4}_{2}\mathrm{He} + \mathrm{n} &\rightarrow \\ {}^{4}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He} &\rightarrow \end{aligned} \tag{9.1}$$

Therefore the step to the next stable nucleus $^{6}{}_{3}$Li can only proceed via the reaction

$$ {}^{2}_{1}\mathrm{H} + {}^{4}_{2}\mathrm{He} \rightarrow {}^{6}_{3}\mathrm{Li} \tag{9.2}$$

which is, however, during the short period when $^{2}{}_{1}$H as well as $^{4}{}_{2}$He coexist, very unlikely and reflects the fact that only very small percentages of Li had been generated during the early stage of the universe.

Therefore all heavy elements, which we can find today in the universe and in particular on the earth (Fig. 9.2) must have been produced by nuclear fusion processes in the central regions of stars or during explosions of stars, because these are the only situations where not only sufficiently high temperatures are reached for the nuclear synthesis (the Coulomb barrier must be overcome), but also sufficiently high densities to allow collions between the nuclei and enough time to form the heavier elements.

We have seen in Sect. 6.x that in stars with masses $M > 2.6\ M_{\odot}$ carbon can be sythesized by the 3α-fusion process (5.39)

$$\begin{aligned} {}^{4}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He} &\rightarrow {}^{8}\mathrm{Be} + \gamma \\ {}^{8}\mathrm{Be} + {}^{4}_{2}\mathrm{He} &\rightarrow {}^{12}\mathrm{C} + \gamma \end{aligned} \tag{9.3}$$

In spite of the short lifetime of the intermediate nucleus $^{8}{}_{4}$Be ($\tau = 2.6 \cdot 10^{-16}$ s) this reaction is in the central region of stars possible, because the favorable conditions for this fusion process are present for many million years while in the first 4 min after the big bang such conditions were realized in the earlyl universe only for a few seconds.

In the inner part of massive stars also the heavier elements up to iron are synthesized. Their production proceeds stepwise over the C-burning (5.40), the neon burning (5.41), the oxygen burning (5.42) and the silicon burning (5.43) up to the formation of iron.

Here the nuclear fusion in the central region of stars ends, because there will be no more energy gain for the fusion of heavier elements. The nuclear binding energy has

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics*, Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4_9

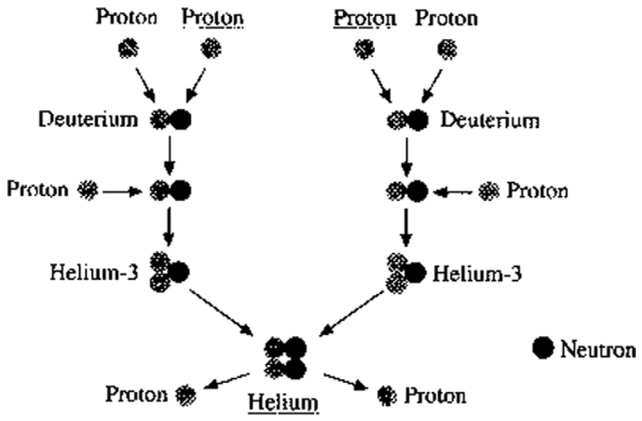


Fig. 9.1 The p–p chain for the production of the light elements deuterium, helium 3 and He 4 in the first minutes of the universe (Wikipedia)

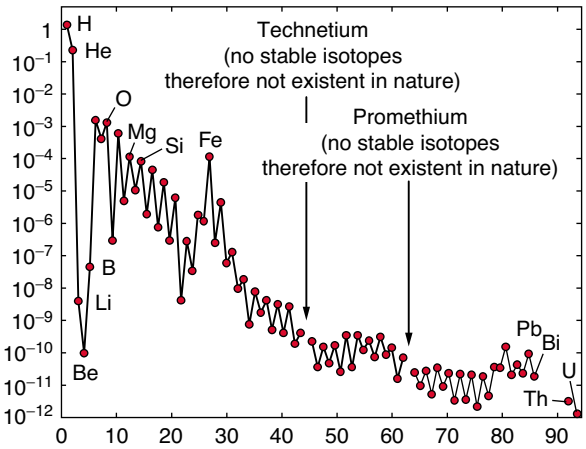


Fig. 9.2 Abundance of the elements in the solar system, normalized onto the most abundant element hydrogen H (from: W. S. Broecker, Labor Erde Springer Heidelberg 1995)

its maximum for the isotope $^{62}_{28}$Ni (Fig. 9.3), Since the isotope $^{62}_{28}$Ni is very rare, the most important heaviest element of nuclear fusion in stars is the iron isotope $^{58}_{26}$Fe.

There is another reaction which allows the production of heavier nuclei: This is the neutron capture

$$z^{A}N + n \rightarrow z^{A+1}N \tag{9.4}$$

where a heavier isotope of the initial nucleus is produced. The neutron is not hindered to penetrate through the Coulomb wall. Since this isotope has a smaller binding energy, it is generally not stable and can for instance pass by ß-decay into a new element:

$$z^{A+1}N \rightarrow {}^{A+1}_{Z+1}N + e^{-} + \bar{\nu}_e \tag{9.5}$$

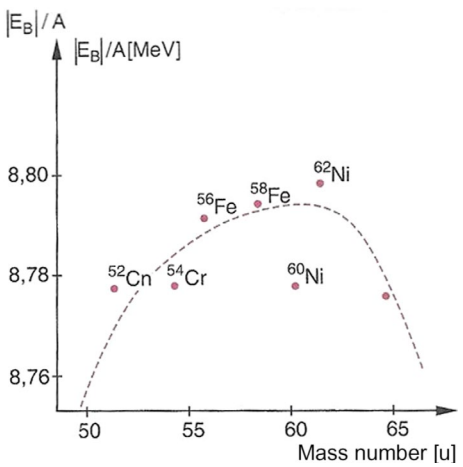


Fig. 9.3 Elements around the maximum binding energy per nucleon

Under the normal fusion conditions in stars generally only a few neutrons are present They can be, for instance, produced by the fusion process

$$6^{12}C + p \rightarrow 7^{13}N + \gamma \rightarrow {}^{13}_{6}C + e^{+} + \nu_e$$

$$^{13}_{6}C + {}^{4}_{2}He \rightarrow {}^{16}_{8}O + n \tag{9.6}$$

Much larger neutron fluxes are generated during the final states of stars, where the star passes through instable phases which are connected with fast contractions and expansions. In such processes shock waves are generated. In the highly compressed shockfronts such high temperatures are present, that heavy nuclei (e.g. iron) which had been already produced in stars during their stable phase, can be split again. This releases neutrons which can initiate the reaction (9.4)

In particular during supernova explosions shock waves with very high amplitudes are generated, which lead to extremely high densities and temperatures. This enhances the processes discussed above.

If stable isotopes are produced by neutron capture they can further capture new neutrons and generate heavier stable elements until an unstable isotope is reached. Its fate depends on the ratio of the probabilities for ß⁻ decay and further neutron capture [1].

Depending on the neutron flux density two different kinds of neutron capture become important:

In the **slow neutron capture** (s-process) the neutron flux is small. In this case the new isotopes formed by the reaction (9.5) can pass into new nuclei by ß⁻-decay before the next neutron capture occurs. The steps for the generation of heavy nuclei then proceed according to the scheme (Fig. 9.4).

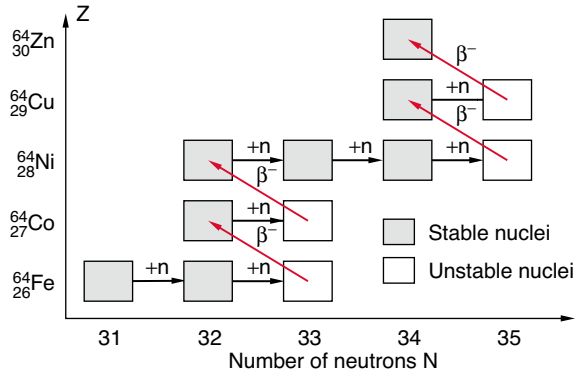


Fig. 9.4 Build-up of the heavy elements by s-processes (slow neutron capture)

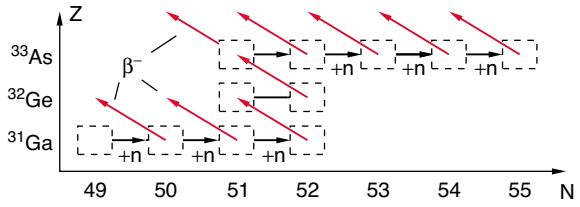


Fig. 9.5 Reaction chain for r-processes (rapid neutron capture)

$$\begin{aligned} {}_{z}^{A}\mathrm{N} &\rightarrow {}_{Z}^{A+1}\mathrm{N} \rightarrow \\ &\rightarrow {}_{Z+1}^{A+2}\mathrm{N} \end{aligned} \tag{9.7}$$

By these s-processes stable nuclei up to mass numbers A = 210 can be produced.

At very high neutron fluxes, which occur during supernova explosions, a second neutron can be already captured, before the nucleus is tranfered by ß− emission into a new nucleus. During **t**hese **r-processes** (rapid neutron capture) the reaction chain proceeds as (Fig. 9.5).

$$\begin{array}{l} {}_{Z}^{A}\mathrm{N} \rightarrow {}_{Z}^{A+1}\mathrm{N} \rightarrow {}_{Z}^{A+2}\mathrm{N} \cdots . \\ \uparrow\,\uparrow\,\uparrow \\ \mathrm{n\,n\,n} \end{array} \tag{9.8}$$

During the short time span of a supernova explosion a chain of neutron-rich nuclei are produced which subsequently pass into stable isotopes by ß⁻ decay [2].

Remark The ß-decay

$${}^{56}\mathrm{Ni} \rightarrow {}_{6.1\mathrm{d}}^{56}\mathrm{Co} \rightarrow {}_{78.8\mathrm{d}}^{56}\mathrm{Fe} \tag{9.9}$$

is, after the emitted neutrinos, the most important energy source for the emission of radiation after the supernova explosion. It is responsible for the prolonged emission of radiation which lasts for several months after the maxiumum of the light curve (see Fig. 7x)

9.2 The Origin and Development of Our Solar System

The fact, that in our solar system heavy elements are found, such as lead or uranium, proves that the matter, our solar system is made of, must have been formed in one or several supernova explosions. This means, that our sun cannot be a star of the first generation, wich had been born in the early stage of our galaxy. It must be a star with medium age.

The age of the earth can be determined with the radioisotope dating (see Vol. 4 Sect. 8.1.6) as $T = 4.5 \cdot 10^9$ years [2]. Since the earth has been formed shortly after the formation of the solar system by gravitational collection of matter from the surrounding of the sun, the age of the solar system must be of the same order of magnitude.

In Sect. 5.2 it was outlined how protostars are formed by the collaps of gas- and dust-clouds, if their densities surpasses the critical Jeans-mass. Such an increase of the density of clouds can be triggered by a supernova explosion, because the ejected gas produces shockwaves in the interstellar medium, which cause a compression and therefore an increase of the gas density [3].

The details of the formation of our solar system are still not completely clarified, but gradually a consens about the correct model is reached, which correctly describes most experimental results.

According to the Jeans-criterion a gas cloud with the mass $M = 1\ M_{\odot}$ collapses if its radius becomes below

$$R \leq G \cdot \frac{\overline{\mu} \cdot m_{\mathrm{H}}}{2kT} \cdot M_{\odot} \tag{9.10}$$

Example

With $\overline{\mu} = 2.5$, $M = M_{\odot} = 2 \cdot 10^{30}$ kg, $T = 10$ K the critical radius becomes $R_c = 5 \cdot 10^{15}$ m $= 3.4 \cdot 10^4$ AU $= 0.17$ pc. This gives the mean density $\overline{\rho} = 4 \cdot 10^{-18}$ kg/m^3 which corresponds to 10^9 molecules per m^3. The collaps time is then according to (5.24) $\tau = 5 \cdot 10^5$ years. ◀

9.2.1 Collaps of the Rotating Gas Cloud

Since our sun has emerged in a spiral arm of our galaxy the gas cloud participated in the rotation of the galaxy. After the start of its collaps, the cloud decoupled from its surroundings but kept its angular momentum.

The angular velocity ω of the rotating galaxy can be obtained from the Oort's relation

$$\begin{aligned} \omega &= \tfrac{1}{R}\left(\tfrac{\mathrm{d}v}{\mathrm{d}R}\right) \cdot R = -(A + B) \\ &= -2\ \mathrm{kms^{-1}kpc^{-1}}, \end{aligned} \tag{9.11}$$

where A and B are the Oort's constants (see Sect. 7.x) Inserting the numerical values of the example above $R = 5{\cdot}10^{15}$ m $A+B = 2$ km/s kpc^{-1} gives $\omega = (2$ km/s$){\cdot}(3{\cdot}10^{19}$ m$)^{-1} = 7{\cdot}10^{-17}$ s^{-1}. This corresponds to the initial rotation period of the gas cloud of

$$T = 2\pi/\omega \approx 10^{17}\text{s} = 3 \cdot 10^9 \text{ years}$$

With these data we obtain the angluar momentum of the gas cloud (which we have assumed as homogenous) before the collaps

$$J = (2/5)M \cdot R^2 \cdot \omega = 1.5 \cdot 10^{45}\text{kgm}^2\text{s}^{-1}$$

Since the angular momentum is preserved during the collaps the angular velocity $\omega(R) \propto 1/R^2$ increases after the collaps to a radius R = 1AU to $\omega = 8 \cdot 10^{-8}\text{s}^{-1}$.

which gives a rotational period of $\tau = 2.5$ years. For this estmation we have assumed that the cloud is after the collaps still homogeneous.

When a rotating cloud collapses, the centrifugal acceleration

$$a_c = \omega^2 \cdot R$$

retards the collaps in the radial directions perpendicular to the angular momentum vector. The initial spherical symmetric mass distribution falls down faster in the direction of the anguar momentum **J** but slower in the radial directions perpendicular to **J.** Therefore with increasing density a flat rotating disc devellops. Since the centrifugal acceleration increases with R, it is less effective for small values of R. Therefore close to the center more mass is accumulated than at the edges of the disc. We expect at the end of the collaps a rotating disc with a bulk at the center, similar to the structure of our milky way, but on a much smaller scale.

We know about the present mass distribution in our solar system that the sun contains more than 99% of the total mass of the system. The by far major part of the mass must have been concentrated during the collaps in the central part (Fig. 9.6).

The angular momentum conservation exhibits, however, some difficulties:

The conservation of the angular momentum

$$J = (2/5)MR^2\omega$$

during the collaps from the radius R_1 to $R_2 = R_\odot$ gives for the angular velocity the relation

$$\omega_1/\omega_2 = (R_1/R_2)^2 \qquad (9.12)$$

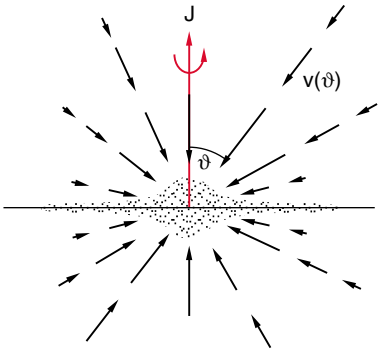


Fig. 9.6 Collapse of a rotating gas cloud

Example

$R_1 = 10^{15}$ m; $R_2 = R_\odot = 7{\cdot}10^8$ m; $\omega_1 = (A+B) = 2$ km/(s·kpac) $\rightarrow \omega_1 = 7{\cdot}10{-}^{17}$ s$^{-1} \rightarrow \omega_2 = (R_1/R_2)^2\omega_1 = 3.5{\cdot}10^{-3}$ s. ◀

This would give a rotation period of $\tau = 0.002$ days. The observed rotation period of the sun (Table 4.10) is, however, about 26 days.

Furthermore the angular momentum of the planets is much larger than that of the sun.

For a contraction of a rotating cloud with mass M and angular momentum J the effective potential (Fig. 9.7) (see Vol. 1, Sect. 1.6) is

$$E_{\text{pot}}^{\text{eff}} = -G \cdot \frac{M(r) \cdot m}{r} + \frac{J}{2mr^2}. \qquad (9.13)$$

It has a minium for $dE_{\text{pot}}^{\text{eff}}/dr = 0$. This gives

$$r_{\min} = \frac{J^2}{2m^2 \cdot G \cdot M}. \qquad (9.14)$$

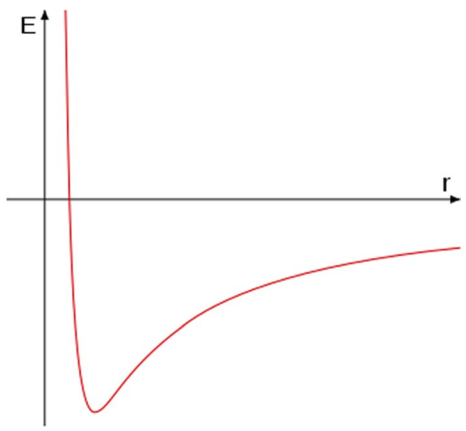


Fig. 9.7 Effective potential

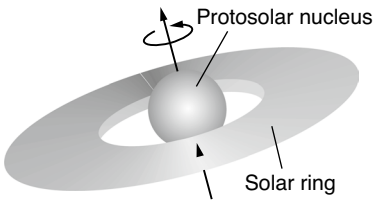


Fig. 9.8 Schematic representation of the early solar system with proto-sun and primitive proto-solar nebula

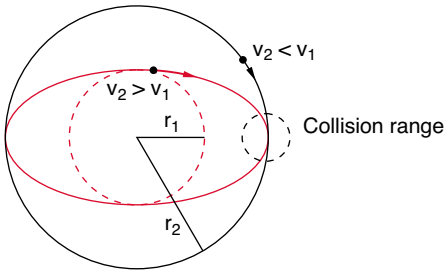


Fig. 9.9 Transport of angular momentum from the central region to outer ranges by collisions

One can therefore expect that during the contraction besides the central mass a solar ring at $r = r_{min}$ is formed (Fig. 9.8). Inserting the numerical values above one obtains for r_{min} a value which is about 0.7 of the Mercure radius.

More detailled model calculations show, that already during the contraction angular momentum is transported from the inner to the outer part of the disc. This can be achieved by collisions. Since the velocity of particles on Kepler orbits decreases with increasing distance from the sun, faster particles on elliptical orbits with a smaller mean distance from the sun can collide with slower particles on orbits with larger distance, and can transfer momentum and angular momentum onto the slower particles, thus increasing their angular momentum (Fig. 9.9).

Another effective mechanism to transport angular momentum is provided by a rotating magnetic field (Fig. 9.10). Such a magnetic field existed already in the galactic plane bevor the collaps of the cloud. As soon as during the collaps the temperature rises the neutral atoms are ionized and the moving charged particles produce their magnetic field which superimposes the original field. The total field participates in the rotation of the contracting cloud and the magnetic field lines reach into the accretion ring. As shown in Fig. 9.10. Since the central part of the magnetic field has a higher rotational velocity, the Lorentz-force of the rotating field can transfer angular momentum onto the ions, which then transfer their angulasr momentum by collisions onto the neutral particles.

This transfer of angular momentum widens the ring, because the radius of the minimum of the effective potential

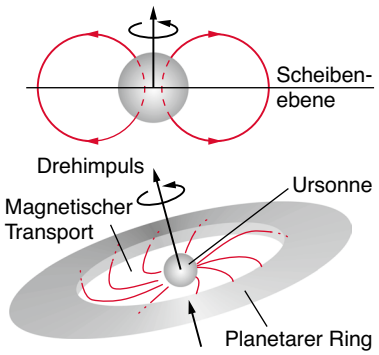


Fig. 9.10 Transport of angular momentum by magnetic interactions

increases with J^2. Simultaneously the radius of the core can further decrease, because it looses angular momentum.

A further possibility of angular momentum transfer can be provided by star winds, i.e. particles emitted by the rotating sun, which take away angular momentum from the sun,

9.2.2 Formation of Planetesimals

During the expansion of the planetary ring at the minimum of the effective potential the potential energy increases. Since the total energy must be preserved, the kinetic energy of the particles decreases. This means that with increasing radius R the temperature of the particles in the ring decreases.

This decrease is different for the gas- and the dust component. In Fig. 9.11 the gas- and dust-temperatures are plotted as a function of the distance R from the sun, based on

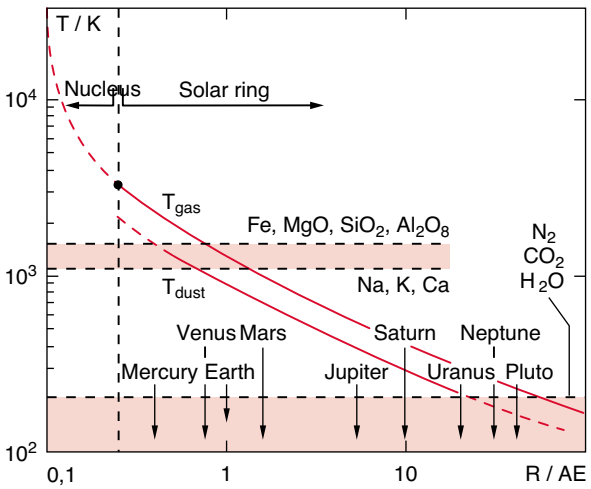


Fig. 9.11 Gas- and dust temperatures at the inner edge of the planetary ring as a function of the distance R from the center (From H. Fahr: Die Bildung des Sonnensystems. Versuch einer Deutung (Phys. Blätt. 37, 142 1981))

calculations by *Fred Hoyle*. The horizontal dashed lines give the condensation temperatures of some elements and gases.

The Figure shows, that at distances of the inner planets the temperatures are still above the condenstation temperatures of gases, but heavy elements, such as Fe, Ni, Na, K, Ca and compounds with high evaporation temperatures could condense. At distances of the outer planets the temperatures are already low enough to allow condensation of gases, such as H_2, N_2, CO_2, H_2O. In Table 9.2 the successive condensation of different elements and compounds are listed.

For the formation of planets not only the gas components are important but also the dust. Model calculations have shown that dust particles with diameters of about 1 mm at distances $R = 10^{11}$ m $= 0.7$ AU could not be heated enough by the proto-sun to melt them. Such particles can be still found today in the ring systems of Jupiter and Saturn (see Sect. 4.x). Such small dust particles accumulate in the thin layer in the central plane of the disc. Since their mass is much larger than that of the atoms, they have at a given temperature a much smaller velocity. They can approach each other with small relative velocities. If their relative distance is small enough, they can suffer inelastic collisions and, due to adhesion forces (el. magn forces), they can stick together. By this way larger particles can collect smaller ones and increase their mass. With increasing mass gravitational attraction plays an increasing role [44].

Such particles do not participate in the angular momentum transport because of their larger mass which reduces the cross section for collisions with much lighter atoms or molecules. They therefore remain in the distance range of the inner planets and could form there small solid bodies, which attract by gravitational forces dust and atoms of their surroundings. Up to diameters of 1 km they are called *Planetesimals*, which are regarded as precursors of the inner planets. This model explains why the inner planets are solid bodies with higher densities ($\overline{\rho} = 4.5 \pm 0.7$ g/cm^3) than the outer planets which consist of gases. There had been much more planetesimals than planets today. By collisions of smaller with larger planetesimals the mass of the larger ones increased further because they could gavitationally bind the smaller ones.

Computer simulations showed that the time between the formation of planetesimales and the origin of proto-planets amounts to about 100.000 years.

It then further took some million years until the proto-planets had accumulate by gravitationao attraction sufficient mass from their surroundings to reach the size of our present inner planets.

Such planetesimals can be still found today as asteroids in the asteroid belt between Mars and Jupiter and also as meteroites which daily fall onto the earth. The analysis of these asteroites gives information about the time of their origin and their chemical composition at this time. Most of these meteroites consist of rocky material. The Chondrites (stone meteroites with small spherical inclusions = chondrens) contain mainly the 4 Elements Si, O, Mg and Fe which constitute 91% of all metals found in meteroites.

Also in the inner planets 4 elements dominates as can be seen for the composition of the earth, listed in ´Table 9.3.

9.3 Separation of Gases and Solid Matter

In the contracting gas cloud hydrogen is mainly present in the form of molecular H_2-gas and in smaller amounts also as methane CH_4 and ammonia NH_3. At a later stage of the collaps the temperatures in the central part rise to such high values, that all molecules dissociate. Molecular gases therefore survive only in the outer parts of the collapsing cloud.

Helium as noble gas does not combine with other elements and therefore does not participate in the condensation. Nearly all of the He-gas either remains in the sun or in the interplanetary space or escapes from the solar system. Small amounts found in the earth atmosphere come from the α-decay of radioactive elements.

The next three elements in the periodic table: Li, Be and B had been synthesized in stars only with a small probability. The small amount of Li, produced in the very early universe, and the share produced in stars by collisions between Helium nuclei and protons was mainly lost by the synthesis of heavier elements. The detection of Li in star spectra indicates that the star is still young. The light elements listed above, are therefore present only in small amounts in the contracting gas cloud. Carbon and Nitrogen combine during the early stage of the collaps with the main constituent hydrogen to methane or ammonia. They survive in the outer part of the collapsing cloud and condense in the outer ranges with lower temperatures, The affinity of oxygen to metals is higher than to hydrogen. Therefore metal oxides are produced which can condense at higer temperatures in the zone of the inner planets, while H_2O remains in the gaseous phase at the high temperatures and can escape as gas into the outer ranges of the solar cloud.

Analogue considerations for all elements in the periodic system leads to results which agree quite well with the observations and confirm the assumption that the contracting cloud as precursor of our solar system had been triggered by a supernova explosion of a star not too far away from the collapsing cloud.

9.4 The Age of Our Solar System

All minerals on earth contain longliving radioactive isotopes which can serve as time keeper, which determine the time since their formation. The method of radioactive dating and its application to the determination of the age of the solar system and the earth will be shortly discussed [5].

The minerals on earth have been at least one times melted since their first formation. During this melting process the volatile radioactive substances have left the mineral and in such cases the radioactive dating gives only the time since the last melting process, but not neccessarily the age of the earth.

The time of crystallization of minerals in chondric meteorites coincides with the early stage of the formation of the inner planets. The structure of the chondrites, in particular the small spheres in chondrites reveal that they had not been melted after their origin. Since the time of cooling from the high temperature phase into the colder solid phase the inner part of a chondric meteorite represents a closed chemical system.

Since the meteorites are the precursors of the inner planets they contain information about the conditions at the time of their formation and about the date of the formation.

The detailled age determination is based on measurements of the abundance of radioactive rubidium ^{87}Rb and of its stable daughter product ^{87}Sr, which appear as trace products in small concentrations (3–10 ppm) in meteorites (Table 9.4). If we choose the time of the formation of the meteorites as $t=0$, and today as $t=t_0$ the number of ^{87}Rb nuclei today is

$$N\left(^{87}\mathrm{Rb}, t_0\right) = N\left(^{87}\mathrm{Rb}, 0\right) \cdot \mathrm{e}^{-t_0/\tau} \qquad (9.15a)$$

With the decay time $\tau = t_{½}/\ln 2 = 7.07 \cdot 10^{10}$ years. For the number of strontium nuclei we obtain

$$N\left(^{87}\mathrm{Sr}, t_0\right) = N\left(^{87}\mathrm{Sr}, 0\right) + \left[N\left(^{87}\mathrm{Rb}, 0\right) - N\left(^{87}\mathrm{Rb}, t_0\right)\right] \qquad (9.15b)$$

In order to get from the ratio of the two isotopes

$$\frac{N\left(^{87}\mathrm{Rb}, t_0\right)}{N\left(^{87}\mathrm{Sr}, t_0\right)} = \frac{\mathrm{e}^{-t_0/\tau}}{\left[N\left(^{87}\mathrm{Sr}, 0\right)/N\left(^{87}\mathrm{Rb}, 0\right)\right] + \left(1 - \mathrm{e}^{-t_0/\tau}\right)} \qquad (9.15c)$$

measured today the time t_0, one has to know the decay time τ and the ratio $N(^{87}Sr,0)/N(^{87}Rb,0)$ at the time of their origin (Figs. 9.12, 9.13 and 9.14).

Here the following method is useful: Besides the unstable radioactive isotopes and the isotopes formed by radioactive decays, there are non-radioactive stable isotopes ^{84}Sr; ^{86}Sr; ^{88}Sr and ^{85}Rb which have not been formed by radioactive decay of other isotopes. Their abundance is therfore constant in time.

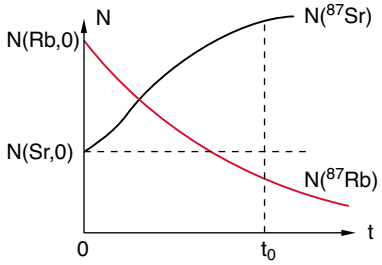


Fig. 9.12 Ratio of the numbers $N(^{87}Rb)/N(^{87}Sr)$ as a function of the age of the meteorite sample

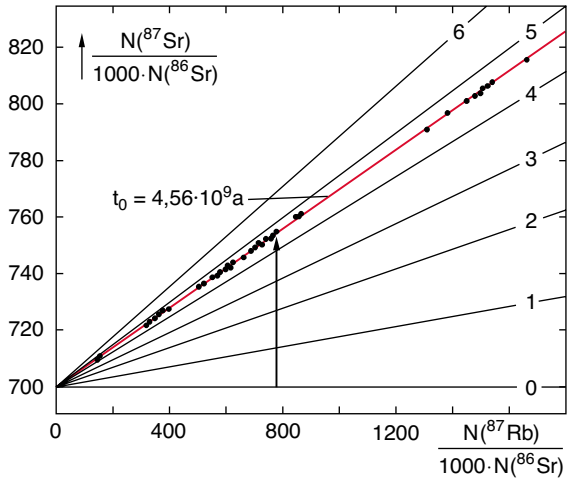


Fig. 9.13 Ratio $N(^{87}Sr, t_0)/(^{86}Sr, t_0)$ plotted against the ratio $N(^{87}Rb, t_0)/(^{86}Sr, t_0)$ for different assumed mineral ages in billions of years (From W. S. Broecker: Labor Erde, Springer Heidelberg 1995)

Now different meteorites with diverse concentrations of Rubidium and Strontium are investigated. The measured ratio $N(^{87}Sr)/N^{86}Sr)$ is plotted against the ratio $N(^{87}Rb)/N(^{86}Sr)$. This gives, according to (9.15b)

$$\left(\frac{N\left(^{87}\mathrm{Sr}\right)}{N\left(^{86}\mathrm{Sr}\right)}\right)_{t_0} = \left(\frac{N\left(^{87}\mathrm{Sr}\right)}{N\left(^{86}\mathrm{Sr}\right)}\right)_{t=0} + \left(\frac{N\left(^{87}\mathrm{Rb}\right)}{N\left(^{86}\mathrm{Sr}\right)}\right)_{t=0}\left(1 - \mathrm{e}^{-t_0/\tau}\right). \qquad (9.17)$$

If the decay time $\tau >> t_0$ (see Table 9.4) we can approximate $1\text{-}e^{t0/\tau} \approx t_0/\tau$ and we obtain the linear relation

$$\left(\frac{N\left(^{87}\mathrm{Sr}\right)}{N\left(^{86}\mathrm{Sr}\right)}\right)_{t_0} = \left(\frac{N\left(^{87}\mathrm{Sr}\right)}{N\left(^{86}\mathrm{Sr}\right)}\right)_{t=0} + \left(\frac{N\left(^{87}\mathrm{Rb}\right)}{N\left(^{86}\mathrm{Sr}\right)}\right)_{t=0} \cdot t_0/\tau, \qquad (9.18)$$

Which is plotted in Fig. 9.15) for different values of t_0 (in billions of years). The comparison with the experimental results gives the age

$$t_0 = (4.56 \pm 0.01) \cdot 10^9 \mathrm{a}$$

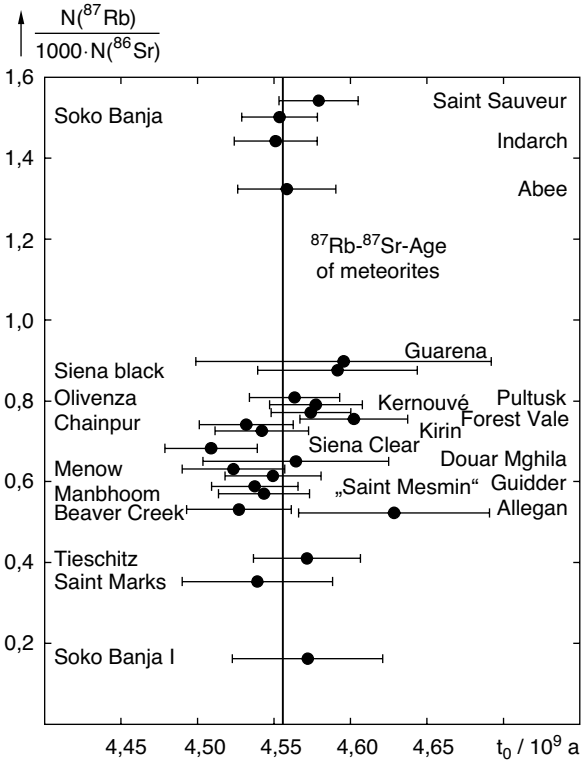


Fig. 9.14 Age determination with the Rubidium-Strontium method for different meteorites with different ratios N(^{87}Rb)/N(^{86}Sr) (From W.S. Broecker: Labor Erde, Springer Heidelberg 1995)

the accuracy of such age determinations is illustrated in Fig. 9.16), which shows age determinations with their error bars based on measured ratios N/^{87}Rb)/N(^{86}Sr) in 19 different meteorites [2]. The accordance of the resuts on completely different samples is impressive.

We Can Therefore Safely Make the Statement that the Earth is About 4.5. Billion years Old.

There are some other radiactive dating methods, for instance the uranium-lead dating or the Pb-Pb-technique, which is a method for dating geological rock samples of materials such as granite or of mateorites yielding the age of the earth.

It can be shortly explained as follows [5, 6] (Fig. 9.13).

There are three stable lead isotopes resulting from the deacy of uranium and thorium (Figs. 9.15 and 9.16). These are ^{206}Pb, ^{207}Pb, and ^{208}Pb while the stable isotope ^{204}Pb is not the endproduct of a radio-active decay chain but is the only non-radiogenic lead isotope (Figs. 9.17 and 9.18).

In the course of time the abundance of the radiogenic lead isotopes increases because of the constant decay rate of the parent isotopes, while the number of the stable ^{204}Pb isotope remains constant. This can be described by the equations

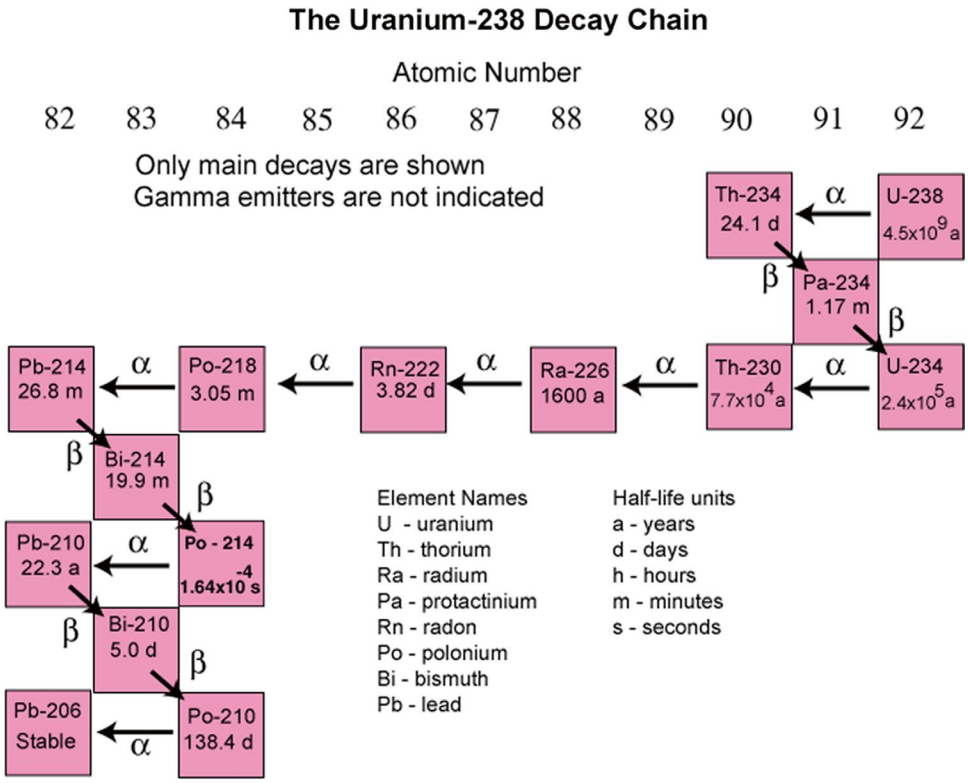


Fig. 9.15 Deacay chain of the Uranium isotope ^{238}U leading to the stable endproduct ^{206}Pb (uranium decay—Bing images)

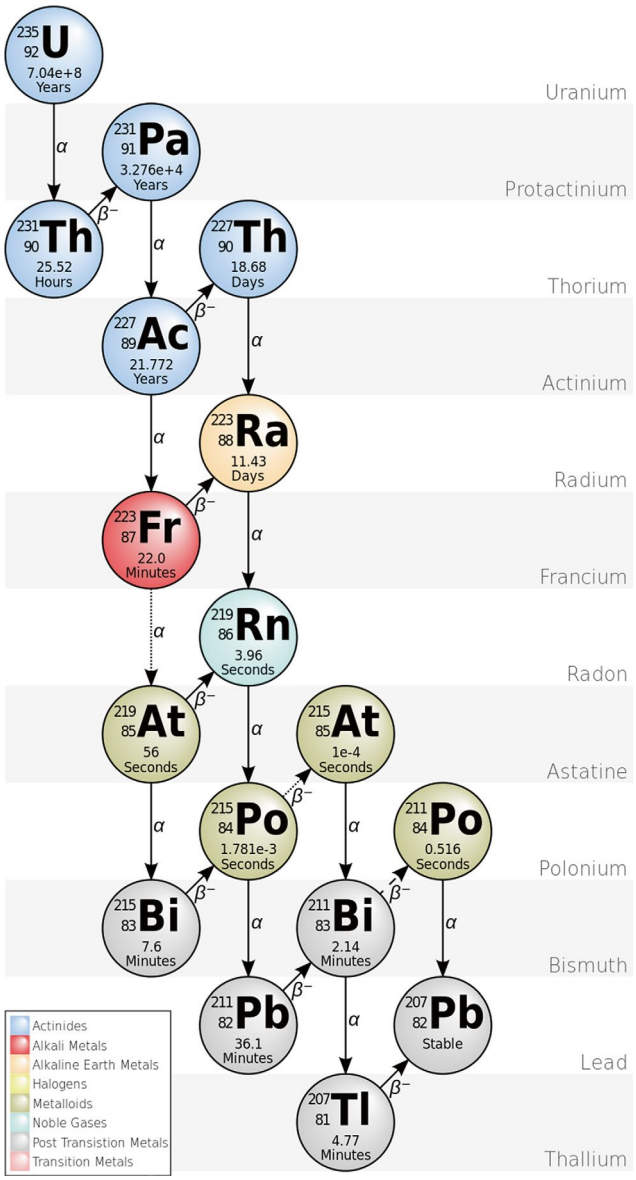


Fig. 9.16 Deacay chain of the uraniumk isotope ^{235}U leading to the stable endproduct ^{207}Pb (Wikipedia)

$$\left(\frac{^{207}\text{Pb}}{^{204}\text{Pb}}\right)_P = \left(\frac{^{207}\text{Pb}}{^{204}\text{Pb}}\right)_I + \left(\frac{^{235}\text{U}}{^{204}\text{Pb}}\right)_P \left(e^{\lambda_{235}t} - 1\right) \quad (9.19a)$$

$$\left(\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right)_P = \left(\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right)_I + \left(\frac{^{238}\text{U}}{^{204}\text{Pb}}\right)_P \left(e^{\lambda_{238}t} - 1\right) \quad (9.19b)$$

where the index p means the present state and I the initial state. The exponent λ gives the known decay constants of the 235 or 238 Uranium isotope. Dividing the first by the second equation gives

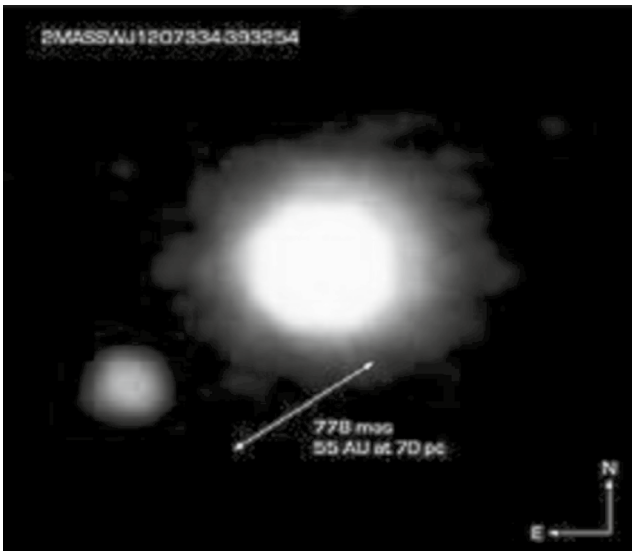


Fig. 9.17 **a** Comparison of the sizes of Jupiter and the exoplanet TrES-4, one of the largest exoplanet discoverd so far **b** the earth compared with the exoplanet Kepler-186 [Aldaron Wikimedia]

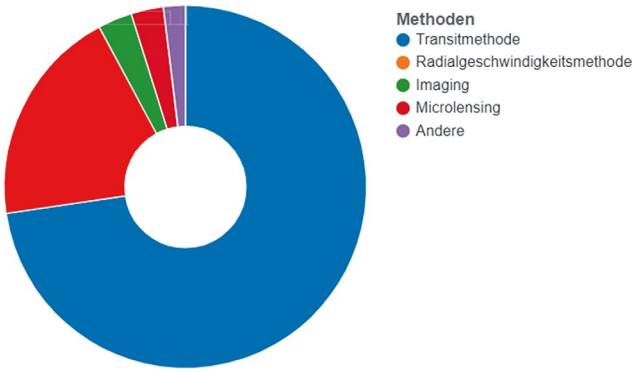


Fig. 9.18 Fraction of exoplanets discovered by the different detection methods [https://de.wikipedia.org/wiki/Extrasolar_Planets_Encyclopaedia]

$$\left[\frac{\left(\frac{^{207}\text{Pb}}{^{204}\text{Pb}}\right)_P - \left(\frac{^{207}\text{Pb}}{^{204}\text{Pb}}\right)_I}{\left(\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right)_P - \left(\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right)_I}\right] = \left(\frac{1}{137.88}\right)\left(\frac{e^{\lambda_{235}t} - 1}{e^{\lambda_{238}t} - 1}\right) \quad (9.19c)$$

where 137.88 is the present ratio ^{138}U/^{135}U.

If the present ratios ^{207}Pb/^{204}Pb and ^{206}Pb/^{204}Pb are measured in a meteorite (which can be regarded as a closed system) where the number of the lead isotope ^{204}Pb does not change with time while that of the radiogenic isotopes ^{207}Pb and ^{206}Pb increases linear with time due to the constant decay rate of the parent isotopes. The extrapolation to a value t_I where (9.19c) is fullfiled, yields the age

of the meteorite and with it also the age of the early solar system [7].

There is another interessant aspect of the measurements of isotope abundances in meteorites. One can get information about the time of the origin of heavy elements produced by the r-process during supernova explosions. This process must have happen before the origin of our solar system. The question is: Was such a supernova explosion really the trigger for the collaps of the gas cloud or did the gas cloud already contain heavy elements from former supernovae explosions.?

One can today simulate by model calculations the conditions during a supernova explosion quite well. The accuracy of such model calculations can be prooved by comparison of the calculated with the measured cross sections for neutron capture. The isotope ratios obtained by such models are

$$R(0) = N\left(^{235}\mathrm{U}\right) : N\left(^{238}\mathrm{U}\right) : N\left(^{232}\mathrm{Th}\right) = 79 : 52.5 : 100. \tag{9.19d}$$

However from measurements in chondrit meteorites one obtains

$$R(t_0) = 0.19 : 26.5 : 100. \tag{9.19e}$$

With the known decay times (Table 9.4) the ratios at $t = t_0 = 4.5$ Billion years ago at the origin of our solar system can be determined as

$$R(t=0) = \left(e^{+t_0/\tau_1} \cdot N\left(^{235}\mathrm{U}, t_0\right)\right) / \left[\left(e^{t_0/\tau_2} \cdot N\left(^{238}\mathrm{U}, t_0\right)\right) : \left(e^{t_0/\tau_3} \cdot N\left(^{232}\mathrm{Th}, t_0\right)\right)\right], \tag{9.19f}$$

This differs from the model calculations which give the ratios (9.19a). If one trusts the model calculations, it follows that the origin of the elements must be dated long before the origin of our solar system.

When we regard as example the measured ratio of the numbers of the two uranium isotopes $N(^{235}U)/N(^{231}U)$ we get for the time of the origin of the elements an estimate of 2 Billion years bevor the origin of our solar system.

We know, that the average rate of supernovae in our galaxy is about 1 per 100 years. Therefore it is probable, that the heavy elements do not originate from a single supernova but possible from several explosions.

The period of the rotation of our galaxy is about 200 million years. Therefore elements which have been produced at different times and therefore at different locations in our galaxy, can have been distributed over larger regions during billions of years, due to the differential rotation of the galaxy and also due to turbulences.

The last supernova in the surroundings of the birth place of our solar system then may have triggered the collaps of the gas cloud as precursor of our solar system [8].

9.5 Other Solar Systems and Exoplanets

During recent years an intense search for other solar systems (planets which orbit around another star) have been performed. Meanwhile more than 4700 extra-solar planets (exo-planets) have been discovered. A list of all exoplanets found up to date together with their properties (mass, temperature and surface characteristics, an be found in [9, 10].

The interesting question is now, whether exoplanets exist with characteristc properties that would allow the development of intelligent life. Up to now, several such habitable planets have been identified [11].

The number of new exoplanets discovered per year are listed in Table 9.1 (Fig. 9.19).

9.5.1 Observation Methods

The difficulty in the observation of extra-solar systems at far distances is the brigth luminosity of the central star and the by many orders of magnitude weaker radiation of the non self- luminous planets. This has the effect, that the small fraction of the radiation of the star, that falls onto the planet and is scattered into the observation telescope, is completely outshined by the radiation of the central star. One has to use special tricks to filter out the weak radiation from the planet.

The oldest observation method which had lead to the discovery of exoplanets is an indirect method. It is based on the fact that star and planet orbit around their common

Table 9.1 List of new exoplanets discovered per year

Number of discovered exoplanets per year [27] as known up to the 12th of February 2021									
1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
1	0	0	0	3	0	0	3	7	0
1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
8	12	23	15	33	26	35	35	35	60
2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
65	93	130	208	144	179	872	170	1541	171
2018	2019	2020	2021						
357	197	233	19						

Entdeckungen nach Methode [27] (12. Februar 2021)

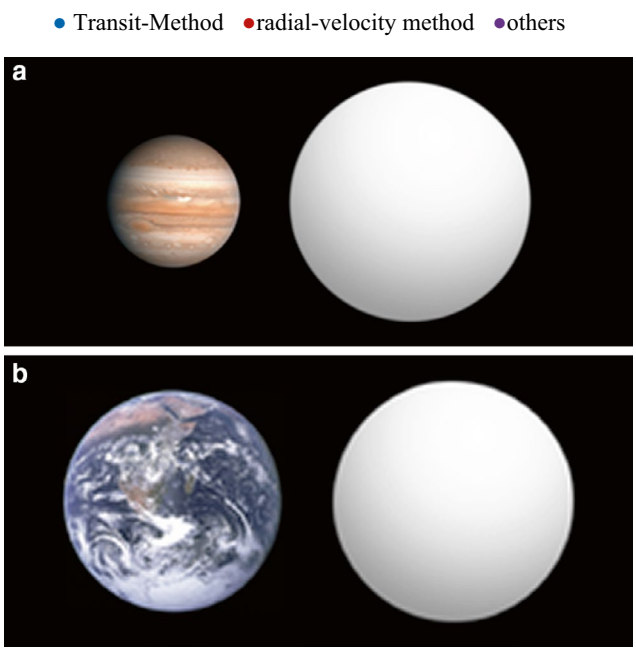


Fig. 9.19 **a** Comparison of the sizes of Jupiter and the exoplanet TrES-4, one of the largest exoplanet discovered so far **b** the earth compared with the exoplanet Kepler-186 (Aldaron Wikimedia)

center of mass (Fig. 5.11). The motion of the central star is measured, i.e. its periodic deviation from a straight line. This allows the determination of the revolution period of the planet and, if the mass of the star is known one can also determine the mass of the planet and its distance from the central star (see Sect. 5.1.2). Only in a few cases the motion of the star can be directly measured. (This became possible with the astrometric Satellite GAIA (see Sect. Yy). Generally its motion is indirectly measured by the Doppler shift of its spectral lines. Here the very large telescope VLT on the mountain Cero Amazonas in Chile, which is under construction, will open new possibilities for more accurate measurements of Doppler shifts.

If more than one planet orbits around the central star, (up to now more than 200 of such multi-planet-sytems have been found), the motion of the star is more complicated. The Fourier-analysis of this motion allows the separation of the different contributions of the different planets and gives their revolution periods, using Keppler's third law. This also allows to calculate their relative distances from the central star. For the correct analysis of the data, one has to know the inclination angle of the planetary plane against the line of sight to the earth.

If the plane lies in the line of sight the planet passes once per revolution period in front of its star and covers part of its surface. This results in a reduction of the received star radiation. From the measured time distance between these radiation minima the revolution period of the planet is immediately obtained. The advantage of this method is, that not the weak light from the planet but the much larger radiation from the star is measured.

In 2004 the first direct observation of an exoplanet has been reported (Fig. 3.19). and in 2005 Astronomers in Jena, Germany, published a photo of an exoplanet orbiting around the young star GQ Lupi, which is only 2 Million years old. This photo was made with infrared radiation with the large telescope of the ESO on the Paranal in Chile. On this planet no intelligent life could have been developed because it is too young.

Up to now more than 4700 exoplanets have been discovered. Most of them are gaseous planets with a small distance from their central star [12].

Only about 10% of them are similar to the earth. However, one has to consider that larger planets are easier to detect. Therefore there might be a larger fraction of earth-like planets.

9.5.2 Conditions for the Development of Intelligent Life

There are severals restrictions for planets, which must be fullfilled in order to enable the development of life:

1. They must be solid bodies (otherwise every body would fall down towards the center).
2. The surface temperature must be within 0–100 °C in order to keep water on its surface, which is vital for the existence of intelligent life.
3. They must have an atmosphere, which smoothens the temperature differences between day and night.
4. They should rotate fast eneough to minimize the temperature difference between the day- and night side.
5. There must be tetravalent atoms, such as Carbon C or Silicon Si, which have enough binding possibilities to form many different complex molecules as the building blocks of life. Since the heavy elements are rare, C and Si are probably the only atoms which have to be regarded.

The extrasolar system therefore must have been started like our own system, by a supernova explosion which provided the necessary amount of heavy atoms. Also the atoms C and Si were not produced in the early phase of the universe some minutes after the big bang, but had been synthesized in the central part of stars.

Water is necessary for the development of life, because the reaction threshold of biological reactions is lowerd by the presence of water and therefore many reactions, necessary for the development of life, could not proceed without water. Many bio-molecules can only exist in a narrow

temperature range. This implies that the orbit of the planet around its central star should not deviate too much from a circular path. Otherwise the temperature would vary during the revolution period. Our earth can serve as example: its orbit lies between Venus (too hot) and Mars (too cold). It is the only planet in our solar system which fullfills all necessary conditions for the development of intelligent life. Planets orbiting around double stars or multi-star systems cannot fullfill these conditions and therefore no intelligent life can develop.

The mass of the planets should be not too small. Otherwise it cannot keep its atmosphere (see mercure). The planet should be old enough to enable sufficient time for the development of life (on earth it took more than a billion years)

These conditions restrict the number of possible planets for the development of intelligent life. Up to now mainly large gas planets have been discovered because they are easier to detect, There should be, however, a sufficient number of habitable earth-like planets as possible candidates for the home of intelligent life. In 2007 such an earth-like planet, orbiting with 3 other planets around the red dwarf star Giese 581, has been discovered which has a surface temperature between 0 and 40 °C and can have therefore water.

The following estimation shows how many earth-like planets can be expected in our galaxy.

Our galaxy contains 10^{11} stars. Only about 30% are born as single stars. From these 30% only about 10% have a sufficient distance to other stars which enables their planets to orbit around their central star unperturbed by external stars. This excludes stars in the central region of our galaxy, because here the star density is too high. Assuming that about 10% of these single stars have a planetary system and furthermore 10% of these stars are sufficiently old, then we come to about $3{\cdot}10^7$ extra solar systems in our galaxy. The probability thát in these planetary systems earth-like planets exist, is difficult to estimate, but is probably smaller than 1%. According to this crude estimation there would be about 10^5 earth-like planets in our milky way, on which probably intelligent life could develop [13].

Many planets have moons. It could be, that on solid moons of gaseous planets life could develop. This would greatly enlarge the possible civilizations in our galaxy.

Another question is, whether such possible civilizations are comparable to ours. The question, most interesting to most people is whether we could ever comunicate with them? This possibility is greatly reduced by two facts. The first one is the large distance to them. Even if an advanced civilization has developed space ships with much larger speeds than we have available, a distance of more than 30 light years represents a nearly insurmountable barrier for the visit of aliens to our planet unless the lifetime of the aliens is much longer than ours. The second fact is the small probability that these civilizations live at the same time as we. Keeping in mind, that our civilization with the ability to build space ships is only about 100 years old compared with a billion years where it could have existed, this drastical decreases the probability of meeting other civilizations.

In order to receive signals from other civilizations, they should not be farther away than 40 light years, because then a signal send by someone from the earth could be answered only after 80 years. This corresponds to the average lifetime of human beings. Within a sphere around the earth with radius of 40 light years about 300 stars exist, where about half of them shows a system of planets but only a small percentage of them might be habitable.

If their are “aliens” they must live farther away and communication with them is very unlikely [13–14].

9.6 The Development of the Earth

For us as habitants of the earth it is, of course particularly interesting to explore the development mechaisms of the earth and their progression in time. We will therefore at the end of this textbook represent the present knowledge we have today about the processes that have finally lead to the structure of our earth [15].

9.6.1 The Separation of Core and Earth Mantle

We have discussed in the previous section that the inner planets have been formed from planetesimals by the gravitational accumulation of dust, rocky fragments and boulders.

We can deduce from seismic investigations (see Sect. 4.2.3) that the earth has a layered structure, which consists of an inner solid Fe–Ni-core, an outer liquid Fe–Ni-core and a mantle, which consists of liquid plasma and a solid crust consisting of solid minerals (Fig. 9.20).

This shows that the gravitational collection of boulders which partly melted the surrounding of the impact location on the earth due to the impact energy of the boulders, must have been followd by a melting process, which had lead to a separation of metal and silicate mineral phases. The molten metal dropped down to the core while the solid minerals stayed in the mantle. That such a separation really took place can be seen in stone-iron meteorites (Fig. 9.21) which had passed through a melting phase, different from the chondric meteorites.

Radioactive dating reveals, that the separation in metal core and silicon mantle took place already in an early stage, even before the earth has got its final mass. Several

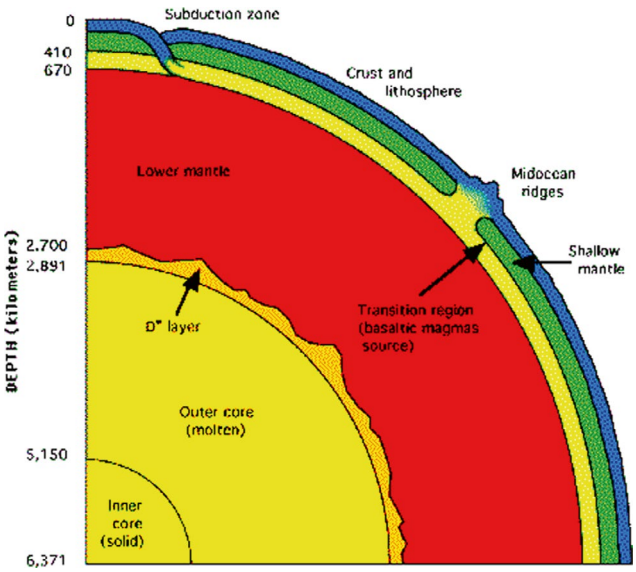


Fig. 9.20 The different layers of the earth [wikipedia]

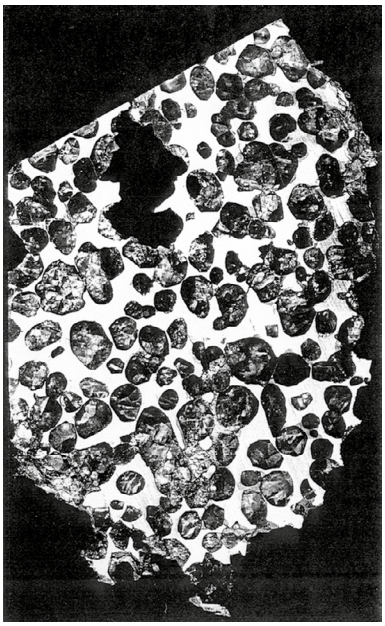


Fig. 9.21 Mineral-iron meteorite (Brenham meteorite). During melting processes the metal- and the mineral phases have been separated (From W. S. Broecker: Labor Erde, Springer Heidelberg 1995)

indications prove that the earth core was formed during the first 100 Million years of the earth history.

One of these indicators is the content of ^{129}Xe in the lava of recent vulcanic eruptions which is higher than in the earth atmosphere. As has been discussed extensively in [16] one can conclude from this fact, that gaseous iodine $^{129}_{53}$ J, which decays with a lifetime of $1.7{\cdot}10^7$a into $_{54}{}^{129}$Xe at least 10^8 years after the formation of the earth during the formation of the liquid core, diffused into the mantle and caused there an enhanced concentration of ^{129}Xe.

A further indication of the early separation of core and mantle is provided by the ratio of the different lead isotopes, which had been produced by the decay of uranium and thorium. This ratio shows that the separation of lead (which dropped in liquid form into the core) and uranium, (which stayed in the mantle in form of uranium minerals because of its high melting temperature) occured already about 10^8 years after the formation of the solar system.

The question is now what caused this separation? Here several models have been developed.

One of these models assumes, that the earth was formed as a mixture of metals and metal oxides in form of boulders with different sizes, ranging from mm to km. The separation, present today has been started by a melting process. The molten metal dropped down to the center of the earth while the silicate minerals with a much higher melting temperature stayed as solids in the earth crust. The melting process needed energy. Three energy sources must be considered:

1. The impact of boulders attracted by gravitational forces between earth and boulders at a time when the earth mass is already much larger then that of the boulders. This energy is produced at the surface. One part of it is transferred to the inner part by heat conduction, another part is sent away by radiation.
2. The gravitational energy increases with increasing mass due to the impact of new boulders. For the mass increase dM/dt per sec the increase of the gravitational energy E_{pot} for a total mass $M(t_0) = M_0$ and a radius $R(t_0) = R_0$ is

$$\frac{dE_{pot}}{dt} = -G \cdot \frac{M(t)}{R_0} \cdot \frac{dM}{dt}. \tag{9.20}$$

The power emtted as radiation, is

$$\frac{dW}{dt} = 4\pi R_0^2 \sigma \cdot T_0^4. \tag{9.21}$$

Without the radiation from the proto-sun a temperature rise would still occur, if the potential energy increase would be larger than the energy loss by emitted radiation, i.e.

$$dE_{pot}/dt > dW/dt$$

Whether this condition can be fullfilled depends on the rate of mass increase dM/dt of the proto-earth by accumulation of small bodies.

Since we do not know this rate we cannot safely say whether this energy supply is sufficient to reach the melting tempearture of metals, in particular of iron.

3. The energy from the decay of radio-active substances inside the earth. This energy increases also with

increasing mass of the earth. because the accumulated small boulders have the same chemical composition as the proto-earth. Here the energy source is uniformly distributed in the whole earth and therefore the losses of this energy by emitted radiation are not so important.

It can be estimated that the energy supplied by the radioactive elements in the early earth (the short lived radioactive elements are meanwhile exhausted) was sufficient to increase the temperature in the inner part of the earth by more than 1.000 K. As soon as the temperature surpassed the melting temperature of a metal the liquid metal drops towards the core of the earth, due to its higher density. This increases the potential gravitational energy, because the matter with the higher density reaches the core and increases its mass.

These three energ sources are by far sufficient to reach the melting temperature even of high melting metals and therefore start the separation of matter in the core, the mantle and the earth crust. According to this model, the radioactive energy serves as trigger of the separation, but the main part of the energy comes from the gravitational energy increase gained by the separation.

Probably the earth owns its present structure to the combination of radio-active and gravitational energy sources. For example in the planet Jupiter the separation process is still going on with the result that Jupiter emits three times more energy than it receives from the sun, but still does not cool down.

9.6.2 The Earth's Crust

The earth crust overcasts the mantel like a thin wrinkled skin. Its mass amounts to less than 1% of the eath mass. It also exists at the ground of the oceans but has there another form and thickness than on the continents. While the continents consist mainly of granite and gneiss, the crust below the oceans is mainly made of basalt (Fig. 9.22). The separation between the earth interior and the crust developed about 100 Million years after the formation of the earth (Fig. 9.23).

In Table 9.1 a comparison is given of the chemical composition in the earth mantle, (consisting of Granit and Basalt) and in the chondric meteorites, where that of the meteorites is similar to that of the boulders the earth is made of. This table also illustrates that the composition of the earth mantle is quite different from that of the crust, which mainly consists of granite. For example the abundance of aluminum Al, sodium Na and potassium K

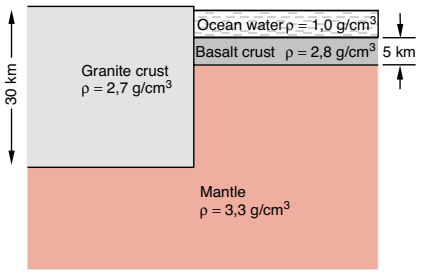


Fig. 9.22 Schematic representation of the granite crust of the continents floating on the viscous magma of the earth mantle, and the basalt crust above the ocean ground (From W. Broecker: Labor Erde Springer Heidelberg 1995)

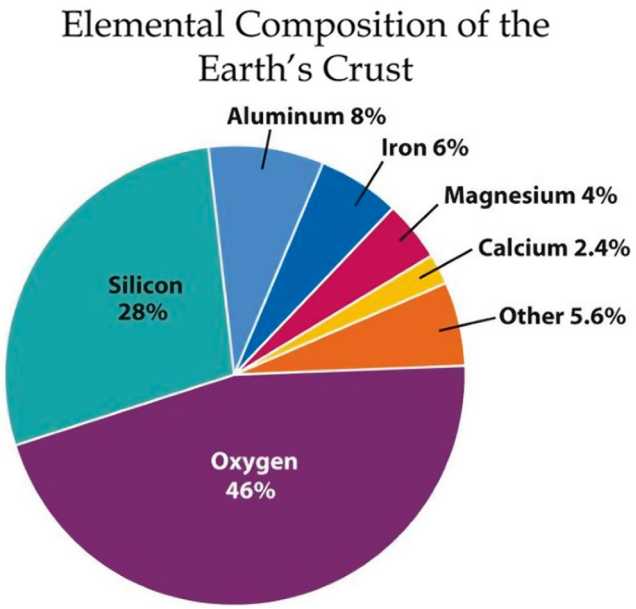


Fig. 9.23 Elemental composition of the earth's crust [Wikipedia]

is higher in the crust, while the mantel contains more Iron Fe and magnesium Mg. The composition of the crust below the oceans with basalt as main constituent is more similar to that of the mantle.

Radio-active dating reveals important differences between the age of the different minerals. The oceanic basalts are generaly younger than 100 Million years, while the granites of the continental crust are up to 3.8 billion years old.

The age of an eruptive mineral is defined as the time since its cristallization but the age of a sediment since its deposition, while the age of metamorphic rocks is defined since its restructering.

The earth crust therefore consists of two quite different parts: the young basalts of the ocean ground and the much older granits of the continents.

9.6.3 Volcanic Activity

Since during volcanic eruptions liquid lava streams from the earth mantle to the surface. It contains minerals with melting temperatures above 1.000 °C and therefore the earth mantle must be at a temperature above 1000 °C even 4.5 billion years after the formation of the earth. The reason for this high temperature is the heat produced by radio-active decay in the earth mantle and the earth core combined with heat conduction from the core to the mantle. Based on seismological measurements we know, however, that the earth mantle consists of solid matter which can become viscous under the high pressure present in the mantle. Driven by temperature gradients, large convection cells develop in which the viscous matter rises to the earth crust. Under the crust it moves sidewards and then sinks down again (Fig. 9.24). Below the crust the temperature is lower than in the mantle but the viscose material can still reach the melting temperature and bubbles of liquid magma can be formed. This bubbles stream through cracks in the crust to the surface (Fig. 9.25).

After cooling down the lava forms part of the crust. Such volcanic eruptions contribute essentially to the formation of the crust. This could be observed some years ago in Iceland where a new island was formed by the volcanic eruption of magma and quite recently on the island Las Palma, where a new peninsula has been formed by the solidified lava.

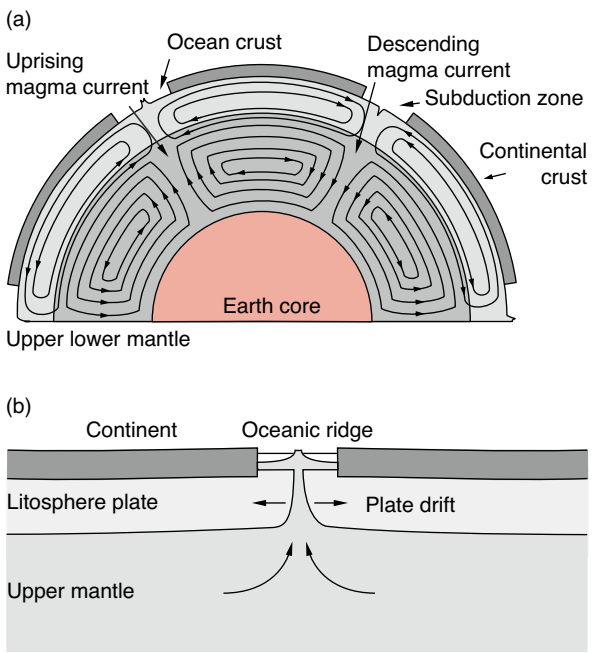


Fig. 9.24 **a** Magma currents in the lower and upper part of the earth mantle. **b** Continental drift, caused by magma penetrating into the earth crust

Fig. 9.25 Eruption of the Mauna Loa on Hawai, the worldwide largest volcano [/www.bing.com/images/search?]

In the small planets like Mercure and also in the asteroids no volcanic action is observed because they have cooled down meanwhile to such low temperatures that no convection currents of magma can be produced. On the Jupiter moon Io,,however, an active volcano has been discovered (Fig. 4.54) because Io is heated by periodic compressions caused by the varying gravitational force by Jupiter on the elliptica orbit of Io around Jupiter.

9.6.4 Formation of the Oceans

Oceans on a planet can be formed only, if several conditions are fullfilled:

- The planet must have collected sufficient amounts of oxygen and hydrogen during its formation in order to form huge quantities of water. In addition the boulders which had formed the earth must have already contained sufficient water.
- This water must have been transported from the inner part of the proto-earth to the surface, or it must have been synthesized on the suface from oxygen und Hydrogen.
- The mean surface temperature of the earth must be between 0 and 100 °C at an atmospheric pressure of 1 bar.

The present fraction of the total water is about 0.5% of the total earth mass. The boulders which formed the earth had

contained probably minerals with enough water to provide the total mass of the present oceans.

During the formation of the earth the temperatures in the inner part of the earth had been above the boiling temperature of water. Therefore H_2O must have existed as water vapor. It could therefore, similar to Xenon, diffuse to the surface of the earth. Above the surface it formed an atmosphere of 80% water vapor, 10% CO_2, 7% H_2S and minor shares of N_2, CO, CH_4 and NH_3. After the cooling of the earth surface (there are now fewer impacts of meteorites), the surface temperature dropped below 100 °C and the water vapor could condense. This started a permanent rain, which lasted for about 50.000 years and filled the depressioins of the earth surface. This was the beginning of the creation of the oceans.

By volcanic actions always new material with a high water content was brought to the surface, increasing the water content. Even above the water surface a water vapor pressure remains which depends on the temperqature. In order to prevent the water vapor from escaping into the space its mean kinetic energy must be smaller than the mean potential energy due to the gravitational attraction by the earth mass. For H_2O-molecules with $T < 100$ °C this condition is fullfilled. However, the H_2O molecules can be dissociated by the ultraviollet radiation of the sun. The ligth H-atoms can indeed escape into the space if their temperature is above 0 °C. Fortunately the temperature in the upper part of the atmosphere is much smaller than 0 °C (see Fig. 4.29), which makes the escape probability very small. The main part of the water vapor which comes from the inner part of the earth to its surface, condenses at the surface and forms the oceans [17]. The total volume of all oceans on earth is about 10^9 km$^3 = 10^{18}$ m^3 corresponing to a mass of about 10^{21} kg ≈ 0.006 m_{earth}. The mean depth of the oceans is 3680 m. The oceans cover about 70% of the total surface of the earth (Fig. 9.26).

Nowadays intense discussions are hold about an increase of the ocean level due to the temperature increase of the atmosphere. This level increase, has two different reasons:

1. The melting of glaciers on a solid surface
2. The increase of the water temperature which leads to an expansion and therefore to an increase of the water volume.

The comparison with the historical increase of the ocean level (Fig. 9.27) shows that this increase for the last 24 thousand years, is mainly due to the melting of glaciers, formed during the last ice-time.

9.6.5 The Origin of the Earth Atmosphere

Our present atmosphere consists of 78.08% Nitrogen N_2, 20.95% of oxygen O_2, 0.93% of Argon A and 0.034% of carbon dioxyd CO_2. The residual 0.006% consist of various

Fig. 9.26 The oceans cover about 70% of the earth surface (Wikipedia)

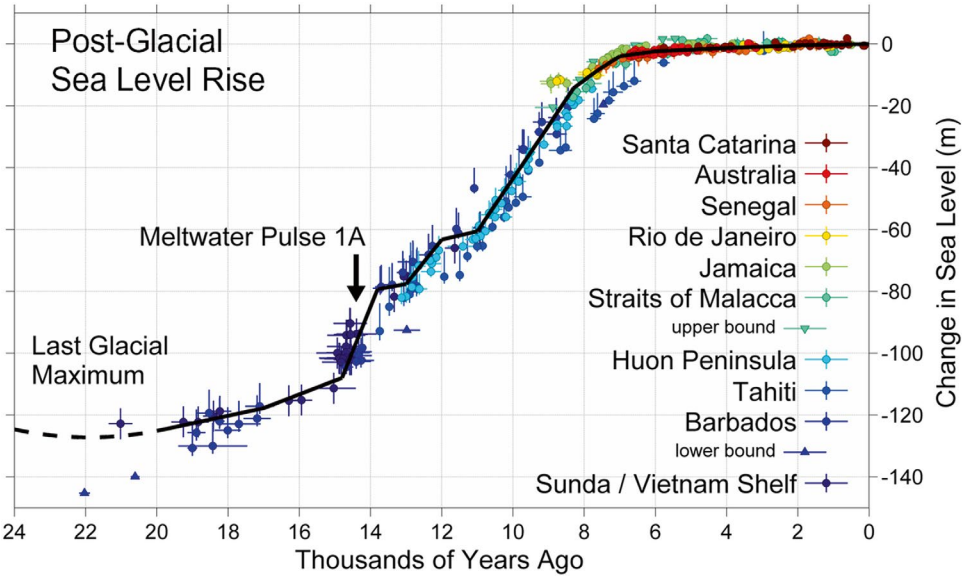


Fig. 9.27 Sea level rise during the last 20.000 years [/www.bing.com/images/search?]

trace gases (Neon, Helium, Krypton, Xenon, Hydrogen) All percentages are given as volume percents (Fig. 9.28).

This Has not Been the Original Atmosphere!

The original atmosphere, which was formed during the early stage of the earth by the gaseous substances of the solar nebula had been blown away by the sun wind of the early sun (T-tauri stage) (Figs. 9.12 and 9.13).

By outgasing of the inner part of the earth, in particular by volcanic action (see Sect. 9.6.3) a new atmosphere was created which consisted mainly of H_2, CH_4, CO, NH_4 and other reducing gas molecules, which did not contain oxygen. Reducing atoms or molecules are those, that easier give away elctrons than accept them. A typical example is hydrogen H_2, which initiates the reaction

$$FeO + H_2 \rightarrow Fe + H_2O \qquad (9.22)$$

and reduces iron oxyd FeO to iron Fe. The oxygen present at the formation of the earth, had been completely exhausted by the reverse reaction

$$2Fe + O_2 \rightarrow 2FeO \qquad (9.23)$$

with iron in the earth core, since the concentration of Fe was larger than that of oxygen.

Since our present atmosphere is oxydising (the oxygen in the atmosphere forms with metals metal oxydes analogue to (9.23)) the question arises which processes had converted the reducing atmosphere into the present oxydising atmosphere and when did this happen?

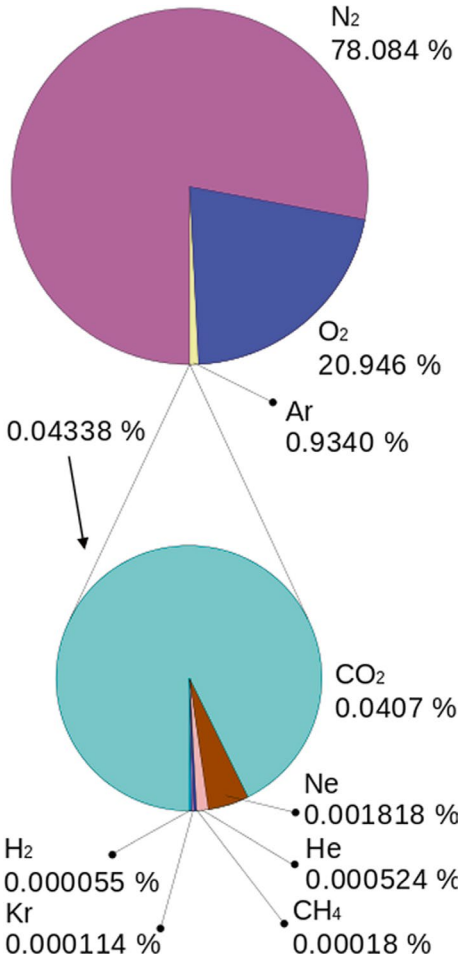


Fig. 9.28 Composition of the present earth atmosphere [Wikimedia]

By the UV-radiation of the sun ($\lambda < 100$ nm) the water vapor could be dissociated

$$H_2O + h \cdot \nu \rightarrow H + HO(\lambda < 220\,\text{nm}), \tag{9.24a}$$

$$HO + h \cdot \nu \rightarrow H + O, \tag{9.24b}$$

and free hydrogen and oxygen atoms are generated. The hydrogen atoms can escape the earth because of their small mass, while the oxygen atoms remain in the atmosphere and can react with each other and a third collision partner M which takes care of energy- and momentum-conservation. Note, that without a collision partner M the two O-atoms cannot combine to form the O_2-molecule, because the kinetic energy of the two O-atoms and their total momentum must be preserved. In the presence of a third collision partner M the molecules O_2 and O_3 can be generated according to the processes

$$O + O + M \rightarrow O_2 + M, \tag{9.25}$$

$$O_2 + O + M \rightarrow O_3 + M \tag{9.26}$$

Ozon O_3 absorbs the sun light already at wavelengths $\lambda < 320$ nm. Under stationary conditions an equilibrium is estabished between the formation and the photo-dissociation of O_3

$$O_3 + h \cdot \nu \rightarrow O_2 + O. \tag{9.27}$$

The important point is now, that the ozone layer, formed in the atmosphere at an altitude between 30 and 50 km, shields the lower atmosphere from the UV radiation of the sun with $\lambda < 320$ nm and thereforer water vapor can no longer by photolysed.

A second important contribution for the formation of oxygen is the photosynthesis of plants, which have been produced at first as algae in the water, where they had been protected from the deadly UV-radiation because of by its absorption by water. Due to the development of Chlorophyl in the plants the photosynthesis

$$6CO_2 + 6H_2O + h \cdot \nu \rightarrow C_6H_{12}O_6 + 6O_2 \tag{9.28}$$

became possible. Here in the reaction centers of Chlorophyll, sugar and oxygen are produced from CO_2 and water, induced by visible photons of the sun radiation.

After the ozone layer had been formed in the higher atmosphere, which absorbed the deadly part of the UV radiation from the sun, plants could safely develop on the continents, because they were now protected against the deadly short wavelenght UV-radiation from the sun.

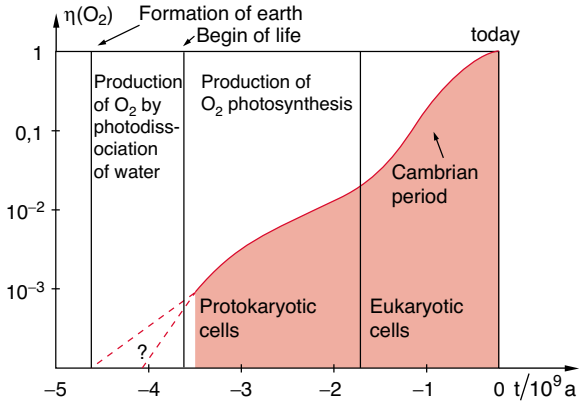


Fig. 9.29 Temporal increase of the oxygen concentration in the atmosphere normalized onto the present value

The oxygen content of our present atmosphere comes from two sources:

- Initially, before the formation of the ozone layer, from the photo-dissociation of water by the short wavelength UV radiation (9.24a) and (9.24b).
- Later, after the establishment of the ozone layer, from the photo-synthesis of plants, induced by visible light.

These two processes have succeeded to convert the reducing atmosphere into an oxydising atmosphere. Only then this has made possibe the development of the variety of life in the water and on the continents, because the necessary energy for the constitution of living cells is provided by the oxydation processes

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \tag{9.29}$$

This "combustion process" is the inversion of the potosynthesis (9.28).

Our present atmosphere is mainly the result of biophysical processes [18]

The increase of the oxygen concentration is illustrated in Fig. 9.29 as a function of the number of years before today.

9.6.6 The Development of Life on Earth

Living Matter is generally defined as such a matter, that can reproduce itself and procreate [19]. The beginning of life on earth took place in the oceans about 3–4 billion years ago. According to our present knowledge it started with chemical reactions in hot chimneys above volcanic sources. In these liquid hot surroundings chemical reactions have produced different forms of primitive life in form of single cell

organisms (protozoom). This may have happened by the development of metabolic processes on the surface of iron-sulfur compounds as can be found in gas-and dust chimneys where hot gases escape from underseas volcanic sources. Nowadays the early life is regarded as an electro-chemical process where under suitable conditions random special processes spontaneously convert in animate matter into a self-producing system. These early forms of life which originated in the hot zones of the oceans without sun light are called *archaea.* They build procariontic cells as bacteria without a cell nucleus. They contain as atomic components carbon C, hydrogen H, oxygen O and some heavy elements. Carbon provides the basis of a molecular skeleton, because it contains 4 open bindings which allows a large variity of spatial molecular structures and furthermore it is very reactive.

In the course of further developments multicellular living structures were created, which moved gradually to the ocean coasts where they were alternately exposed to a water surroundings and to air, due to the tides. Their energy supply changed slowly from the anerobic process to oxygen burning. After the atmosphere had contain sufficient oxygen and ozone, life could further develop on the continents, because it was now safely protected against the deadly UV radiation [20]. Since the formation of ordered structures decreases the entropy, it must increase outside the living organisms, because the total entropy in a closed system must always increase with time. Living organisms do not represent closed systems, because they have to import energy from their surroundings in order to maintain their living conditions.

Remark There exist so called anaerobic bacteria, which can exist without oxygen. Examples are the sulfur bacteria. They use as energy source the reactions.

$$4H_2S + CO_2 \rightarrow CH_4 + 2H_2O + 4S \qquad (9.30)$$

and

$$12H_2S + 6CO_{23} + h \cdot \nu \rightarrow C_6H_{12}O_6 + 12S + 6H_2O \qquad (9.31)$$

where H_2S is emitted by volcanos, and CO_2 molecules are used as starting basis.

It is interesting, that the development of life has itself contributed to create enviromental conditions which are favorable for living organisms.

Nowadays we have to take care, that these favorable conditions are not destroyed by careless and stupid actions, which may have long term effects on our atmosphere and the earth temperature, where the temperature increaseits may no longer be favorable for living organisms [21].

Summary

- The 4 lightest elements have been formed within the first 4 min after the big bang of the universe. These are ${}_1{}^1H$, $D = {}_1{}^2H$, ${}_2{}^4He$ and to a small amount ${}_3{}^7Li$. All other elements up to iron have been produced by nuclear fusion processes in the central part of stars. Heavier elements above iron were formed in the hot regions of supernovae-explosions by neutron capture.
- Our solar system has been formed by contraction of a rotating cold cloud of molecules and dust, probably triggered by a nearby supernova explosion.
- The formation of the solid inner planets and the gaseous outer palanets can be explained by the radial temperature dependence T(R) in the rotating and contracting solar nebula.
- The angular momentum must have been transferred from the central to the outer part of the contracting cloud. Therefore the planets carry about 99% of the total angular momentum, the sun only 1%, while the mass distribution is just opposite: The sun contains about 99% of the total mass of the solar system, all planets together only 1%.
- During the formation of the earth several energy sources could increase the temperature above the melting temperature of some metals. The liquid metals dropped down into the central region of the earth. This resulted in the separation of a Fe–Ni-core and a mantle which contains mainly silicate minerals.
- The oceans could emerge because water, contained in minerals could escape as vapor after the minerals were heated up by the incresing temperature. The saturated atmosphere cooled down and delivered for about 100.000 years heavy rain, which filled the oceans.

Problems

9.1 Show, that a rotating gaseous mass has a minimum of the potential energy, and prove the dependence of r_{min} on the angular velocity ω.

9.2 Calculate the increase Δh of the ocean height h per degree temperature increase for an average ocean depth of 3800 m, when the volume expansion coefficient of water at $T = 15^0$ C is. $\alpha = \Delta V/V = 1.5 \cdot 10^{-4}/K; \rightarrow \Delta h/h = (1/3) \cdot \Delta V/V = 5 \cdot 10^{-5}/K$. $\Delta A/A = 10^{-4}\,{}^0C$.

9.3 Calculate the change of the potential energy, if 10^9 kg of material with density $\rho = 10$ kg/dm^3 drop in the early earth ($M = 5.10^{24}$ kg) from $r = 6000$ km to $r = 1000$ km, if the density $\rho(r)$ follows the function $\rho(r) = \rho_0(1\text{-}r/R)$.

References

1. Andrew McWilliam, Michael Rauch: *Origin and evolution of the elements*. Cambridge Univ. Pr., Cambridge 2004, ISBN 0-521-75578-6. http://courses.washington.edu/bangblue/Cowan-Nucleosynth-PhysTod04.pdf:Claus E. Rolfs, William S. Rodney: *Cauldrons in the Cosmos: Nuclear Astrophysics (Theoretical Astrophysics Series)*. Univ. of Chicago Pr., Chicago 1988, ISBN 0-226-72456-5. L.H. Ahrends: Origin and Distribution of Elements (Elsevier Amsterdam 2015) A.S. Everest: Origin of Chemical Elements. (Taylor and Francis 1957)
2. Age of Earth – Wikipedia
3. [1207.4993] Supernova-Triggered Molecular Cloud Core Collapse and the Rayleigh-Taylor Fingers that Polluted the Solar Nebula (arxiv.org)
4. E. Krügel An Introduction to the Physics of Interstellar Dust. (Taylor and Francis, Boca Raton 2008
5. Radiometric dating - Wikipedia Lead–lead dating - Wikipedia.
6. https://www2.lbl.gov/abc/wallchart/chapters/13/4.html
7. W.S. Broecker, Charles H. Langmuir: *How to Build a Habitable Planet. The Story of Earth from the Big Bang to Humankind*. Princeton University Press, Princeton New Jersey, USA 2012
8. A.P. Boss, S. Iipatov, F.A. Myhill: Triggering presolar cloud collaps (Meteorites and Planetary Science August 2007)
9. Hans Deeg, Juan Antonio Belmonte, Antonio Aparicio (Hrsg.): *Extrasolar planets*. Cambridge University Press, Cambridge 2008, ISBN 978-0-521-86808-2.
10. Rudolf Dvorak (Hrsg.): *Extrasolar planets. Formation, detection and dynamics*. Wiley-VCH-Verlag, Weinheim 2008, ISBN 978-3-527-40671-5.
11. John W. Mason (Hrsg.): *Exoplanets. Detection, formation, properties, habitability*. Springer, Berlin. 2008 https://en.wikipedia.org/wiki/List_of_potentially_habitable_exoplanet
12. Exoplanet – Wikipedia https://de.wikipedia.org/wiki/Exoplanet
13. Peter Umschneider: Intelligent Life in the Universe. Principles and Requirements Behind Its Emergence. (Springer, Heidelberg 2013)
14. I. S. Shklovskii, Carl Sagan: Intelligent Life in the Universe (Emerson Adams 1999)
15. S. Mueller: The structure of the earth's Crust (Elsevier, Amsterdam 2013) Sidney P. Clark: The Structure of the Earth (Bookbarn International 1972)
16. Structure of Earth – Wikipedia
17. Formation of the Oceans (washington.edu) Oceans. Formation of the Oceans. Facts about the Oceans (lubopitko-bg.com)
18. M.Y. Budyko:, A.B. Ronov, A.L.Yanshin: History of the earth's Atmosphere (Springer Heidelberg 1987) https://en.wikipedia.org/wiki/History_of_Earth#:~:text=Earth%20formed%20around
19. Life - Wikipedia life | Definition, Origin, Evolution, Diversity, & Facts | Britannica What Is Life – Wikipedia
20. Steve Jenkins. Life on Earth: The story of Evolution (HMH-books 2020) Nathan Keigley: The History of Life on Earth: The story of evolution on our planet (Austin Macaulty Publ. 2018)
21. Environmental science - Wikipedia Environmental Science: Towards A Sustainable Future: 13th Edition (English Edition) (Pearson Publication 2021)

Solutions of Problems

Chapter 2

2.1 We regard the spherical triangle in Fig. 10.1, where t is the hour angle, A is the azimuth, h is the height, δ = declination, φ = latitude.

With the cosine law

$$\begin{aligned}&\cos\left(90^\circ-\delta\right)=\cos\left(90^\circ-h\right)\cdot\cos\left(90^\circ-\varphi\right)\\&+\sin\left(90^\circ-h\right)\cdot\sin\left(90^\circ-\varphi\right)\cdot\cos\left(180^\circ-A\right)\\&\Rightarrow\sin\delta=\sinh\cdot\sin\varphi-\cosh\cdot\cos\varphi\cdot\cos A.\end{aligned}\tag{10.1}$$

From the sinus law it follows:

$$\begin{aligned}&\frac{\sin(180^\circ-A)}{\sin t}=\frac{\sin(90^\circ-\delta)}{\sin(90^\circ-h)}\\&\Rightarrow\frac{\sin A}{\sin t}=\frac{\cos\delta}{\cosh}\Rightarrow\sin t=\frac{\sin A\cdot\cosh}{\cos\delta}.\\&\cos\tau\quad=-\tan\delta\cdot\tan\varphi,\\&\tau\quad=\text{diurnal semi-arc},\\&\delta\quad=-23.5^\circ\Rightarrow\tau\approx4\text{ h},\\&\delta\quad=+23.5^\circ\Rightarrow\tau\approx8\text{ h}.\end{aligned}\tag{10.2}$$

With $\cos\delta=\sqrt{1-\sin^2\delta}$ the quantity δ can be expressed with (10.2) by h, A and φ.

In an analogue way the transformation from the equator system into the horizon-system is:

$$\begin{aligned}&\cos\left(90^\circ-h\right)=\cos\left(90^\circ-\delta\right)\cdot\cos\left(90^\circ-\varphi\right)\\&+\sin\left(90^\circ-\delta\right)\cdot\sin\left(90^\circ-\varphi\right)\cdot\cos t\\&\Rightarrow\sin h=\sin\delta\cdot\sin\varphi+\cos\delta\cdot\cos\varphi\cdot\cos t\end{aligned}\tag{10.3}$$

and from (10.2) it follows:

2.2 The sun orbit lies in the ecliptic. The declination of the sun varies during the year between $\delta=+23.5^\circ$ at June 21st and und $\delta=-23.5^\circ$ at December 21. From the solution of Problem 2.1, Eq. 2.3 it follows for $\delta=0$ (sunrise, resp. sunset, for the time $t=t_0$)

$$\begin{aligned}\cos\delta&=\frac{\sin A\cdot\cosh}{\sin t}=\frac{\sin A\cdot\sqrt{1-\sin^2h}}{\sin t}\\&\Rightarrow\sin A=\sin t/\sqrt{1-\sin^2h}.\end{aligned}\tag{10.4}$$

$$\begin{aligned}0\quad&=\sin\delta\sin\varphi+\cos\delta\cos\varphi\cos t_0\\\Rightarrow\cos t_0\quad&=-\tan\delta\cdot\tan\varphi.\end{aligned}\tag{10.5}$$

For the latitude $\varphi=50^\circ$ it follows for $\delta=+23.5^\circ$

$$\begin{aligned}\cos t_0\quad&=-0.5182=-\cos(180^\circ-t_0)\\\Rightarrow t_0\;&=180^\circ-58.79^\circ=121.2^\circ\approxeq8^{\mathrm{h}}\,5^{\mathrm{m}},\end{aligned}\tag{10.6}$$

because 15° correspond to 1 h. For $\delta=-23.5^\circ$ is $t_0=58.8^\circ\approxeq3^{\mathrm{h}}\,55^{\mathrm{m}}$. Since for the meridian is t = 0 the time t_0 gives half of the diurnal arc. The diurnal arc $\tau=2t_0$ of the sun for the latitude $\varphi=50^\circ$ has a minimum $\tau_{\min}=7^{\mathrm{h}}\,50^{\mathrm{m}}$ and a maximum $\tau_{\max}=16^{\mathrm{h}}\,10^{\mathrm{m}}$.

Due to the refraction in the earth atmosphere, the final diameter of the sun and the changes of the sun coordinates can be larger by about 10 min.

2.3 According to (2.11) the parallaxes of the stars are $\pi_1=1/3''$ and $\pi_1=1/6''$. The wanted relative parallax, which gives the semi-diameter of the apparent elliptical relative motion is then (Fig. 10.2)

$$\pi_{\mathrm{rel}}=\pi_1-\pi_2=\frac{1''}{3}-\frac{1''}{6}=\frac{1''}{6}.\tag{10.7}$$

2.4 The escape velocity v of a body with mass m from a celestial body with mass M and radius R is determined by the energy conservation

$$\left(\frac{m}{2}\right)v^2=G\cdot\frac{m\cdot M}{R}\Rightarrow v=\sqrt{\frac{2G\cdot M}{R}}.\tag{10.8}$$

(a) Venus: $M=4.87\cdot10^{24}\,\mathrm{kg}, R=6.05\cdot10^6\,\mathrm{m}$
$\Rightarrow v=10.3\,\mathrm{km/s}.$

(b) Jupiter: $M=1.9\cdot10^{27}\,\mathrm{kg}, R=7.14\cdot10^7\,\mathrm{m}$
$\Rightarrow v=60.0\,\mathrm{km/s}.$

(c) Earth moon: $M=7.25\cdot10^{22}\,\mathrm{kg}, R=1.73\cdot10^6\,\mathrm{m}$
$\Rightarrow v=2.4\,\mathrm{km/s}.$

(d) Sun: $M=1.989\cdot10^{30}\,\mathrm{kg}, R=6.96\cdot10^8\,\mathrm{m}$
$\Rightarrow v=618\,\mathrm{km/s}.$

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics,* Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4

Chapter 3

3.1

(a) The diameter of the central diffraction maximum on the moon is

$$\begin{aligned} d &= 2 \cdot 1.2 \cdot (\lambda/D) \cdot r \\ &= 2.4 \cdot \tfrac{5 \cdot 10^{-7}}{0.75} \cdot 3.84 \cdot 10^8 \mathrm{m} \\ &= 6 \cdot 10^2 \mathrm{m}. \end{aligned} \tag{10.9}$$

(b) the actual illuminated area on the moon is

$$\begin{aligned} A_1 &= \pi \, (r_{\mathrm{Mond}} \cdot \sin\alpha/2)^2 \\ &= \pi \left(3.84 \cdot 10^8 \mathrm{m} \cdot \sin\tfrac{3}{2 \cdot 3600}\right)^2 \\ &= 24.5 \cdot 10^6 \mathrm{m}^2. \end{aligned} \tag{10.10}$$

The fraction of the photons falling onto the reflector is

$$\begin{aligned} \eta_1 &= \tfrac{0.4 \times 0.4\,\mathrm{m}^2}{A_1} \\ &= \tfrac{0.16}{24.5 \cdot 10^6} = 6.5 \cdot 10^{-9} \end{aligned} \tag{10.11}$$

(c) The area on the earth illuminated by the reflected photons is

$$\begin{aligned} A_2 &= \pi \cdot \left(3.84 \cdot 10^8 \mathrm{m} \cdot \sin\tfrac{4}{2 \cdot 3600}\right)^2 \\ &= 44 \cdot 10^6 \mathrm{m}^2. \end{aligned} \tag{10.12}$$

The fraction of photons reaching the telescope is

$$\eta_2 = \frac{\pi \cdot (0.75)^2/4}{A_2} = 1.0 \cdot 10^{-8}. \tag{10.13}$$

(d) The energy of the laser pulse is

$$E = P \cdot \Delta t = 0.15\,\mathrm{J}.$$

The energy of the photon is

$$h \cdot \nu = h \cdot c/\lambda = 4 \cdot 10^{-19}\mathrm{J}.$$

Therefore, the number of photons per pulse is

$$N = \frac{E}{h \cdot \nu} = 4 \cdot 10^{17}.$$

Back into the telescopecomes the fraction

$$N_{\mathrm{refl}} = N \cdot \eta_1 \cdot \eta_2 = 25.$$

Due to scattering and absorption in the atmosphere and losses in the telescope actual only 1 photon per pulse is detected.

(e) The mean optical path $\Delta s = \overline{n} \cdot \overline{H}$ through the atmosphere is wit $\overline{n} = 1.00035, \overline{H} \approx 10^4\mathrm{m}$.

The additional pathlength of the laser pulse is then.
$\Delta s - \Delta s_0 = 3.5\,\mathrm{m}$. A change of 1% of Δs gives a change of the path length of 3.5 cm.

3.2 The minimum distance between earth and Venus is (assuming circular orbits with $r_1 = 0.723$ AU, $r_2 = 1$ AU)

$$\Delta s = r_2 - r_1 = 0.277\ \mathrm{AU}.$$

After 30 days Venus with a period of $T_1 = 224.7$ days has proceeded by the angle.

$\alpha_1 = \frac{30}{224.7} \cdot 360° = 48°$.

For the earth is

$$\alpha_2 = \frac{30}{365.25} \cdot 360° = 29.57°.$$

Application of the cosine relation (Fig. 10.3) on the triangle $S'V'E'$ gives:

$$\begin{aligned} \Delta s^2 &= r_1^2 + r_2^2 - 2 r_1 r_2 \cos(\alpha_1 - \alpha_2)^{1/2}\mathrm{AU} \\ &= 0.388\mathrm{AU}. \end{aligned} \tag{10.14}$$

The run time for the radar signal is

$$\Delta T_{2w} = \Delta T_1 \cdot \frac{0.388}{0.277} = 387\,\mathrm{s}.$$

3.3 If a parallel light beam passes through the atmosphere, the optical paths

$$\begin{aligned} s_1 &= (r + \mathrm{d}r) \cdot n(r + \mathrm{d}r) \cdot \mathrm{d}\varphi \text{ and} \\ s_2 &= r \cdot n(r) \cdot \mathrm{d}\varphi \end{aligned}$$

of the different components of the light beam have to be equal, because the phase fronts must be always perpendicular to the propagation direction. Therefore, it follows:

$$\begin{aligned} (r + \mathrm{d}r) \cdot n(r + \mathrm{d}r) &= r \cdot n(r) \\ (r + \mathrm{d}r) \cdot n(n + \mathrm{d}n) &= r \cdot n \\ \Rightarrow r &= -\tfrac{n}{\mathrm{d}n/\mathrm{d}r}. \end{aligned}$$

On the pathlength $\mathrm{d}s = r \cdot \mathrm{d}\varphi$ the light experiences a deflection

$$\mathrm{d}\varphi = \frac{1}{n} \cdot \frac{\mathrm{d}n}{\mathrm{d}r} \cdot \mathrm{d}s.$$

With the apparent zenith distance ξ_a the light path through the atmosphere is.

$$\Delta s = a_1 \cdot \tan\zeta_a$$

where the factor a depends on the height h and the radial dependence n(h) of the refractive index. With $\Delta n = (n(0) - 1)$ the deflection angle becomes $\Delta\varphi = \varrho = \xi_r - \xi_a$: (Fig. 10.4)

$$\begin{aligned} \varrho &= \tfrac{1}{n \cdot \Delta r} \cdot (n(0) - 1) \cdot a_1 \cdot \tan\zeta_a \\ &= a \cdot (n(0) - 1) \cdot \tan\zeta_a. \end{aligned} \tag{10.15}$$

3.4 According to equ. (2.x) is

$$m - M = 5 \cdot \log(r/10\,\mathrm{pc})$$

$$\Rightarrow r = 10\,\text{pc} \cdot 10^{(m-M)/5}$$
$$\Rightarrow r_1 = 10 \cdot 10^{-2.5}\,\text{pc} = 4.0\,\text{pc}$$
$$r_2 = 10 \cdot 10^{0.2}\,\text{pc} = 15.8\,\text{pc}\,.$$

3.5 $L = 4\pi R^2 \sigma \cdot T^4$

$$\Rightarrow T = \left(\frac{L}{4\pi R^2 \sigma}\right)^{1/4} = \left(\frac{10^{27}}{4\pi \cdot 7^2 \cdot 10^{16} \cdot 5.67 \cdot 10^{-8}}\right)^{1/4} \text{K} = 7300\,\text{K}. \qquad (10.16)$$

3.6

$$r = 10\,\text{pc} \cdot 10^{(m-M)/5} = 10 \cdot 10^{(-0.71+4.6)/5}\text{pc} = 60\,\text{pc}.$$

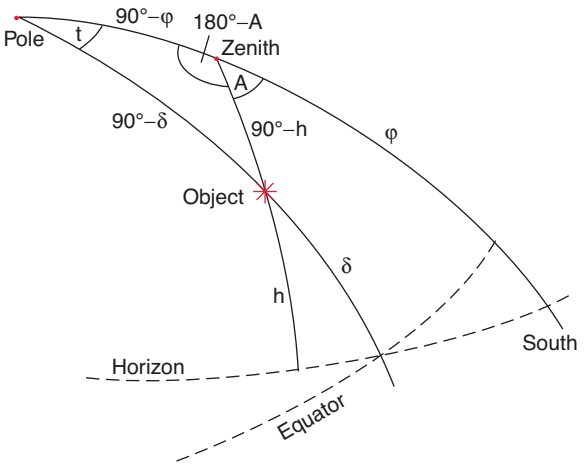


Fig. 10.1 To solution 2.1

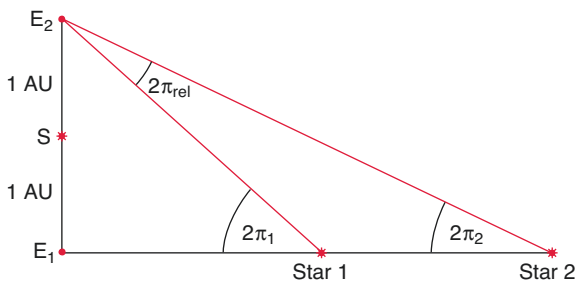


Fig. 10.2 To solution 2.3

Chapter 4

4.1 The angular momentum of the comet with mass m with reference to the sun is

$\boldsymbol{L} = m \cdot (\boldsymbol{r} \times \boldsymbol{v}) = \text{const.}$.
For $r = r_{\min}$ and $r = r_{\max}$ is $v \perp r$.

$$\Rightarrow r_{\min} \cdot v_{\max} = r_{\max} \cdot v_{\min} = C$$
$$= 0.5 \cdot 1.5 \cdot 10^{11} \cdot 5 \cdot 10^4 \text{m}^2/\text{s}$$
$$= 3.75 \cdot 10^{15} \text{m}^2/\text{s} \qquad (10.17)$$
$$\Rightarrow r_{\max} = \frac{C}{v_{\min}}.$$

Energy conservation demandes

$$E = E_{\text{kin}} + E_{\text{pot}} = \frac{m}{2}v^2 - G \cdot \frac{m \cdot M_\odot}{r} = \text{const.} \qquad (10.18)$$

$$\Rightarrow \frac{1}{2}v_{\min}^2 - G \cdot \frac{M_\odot}{r_{\max}} = \frac{1}{2}v_{\max}^2 - \frac{GM_\odot}{r_{\min}}. \qquad (10.19)$$

Inserting (10.17) into (10.19) and multiplying with $r_{\max}^2$ gives:

$$\frac{1}{2}C^2 - GM_\odot r_{\max} = \frac{1}{2}v_{\max}^2 r_{\max}^2 - \frac{GM_\odot}{r_{\min}}r_{\max}^2 \qquad (10.20)$$

(a) It follows

$$\Rightarrow r_{\max} = \frac{GM_\odot}{v_{\max}^2 - 2GM_\odot/r_{\min}} \pm \sqrt{\left(\frac{GM_\odot}{v_{\max}^2 - 2GM_\odot/r_{\min}}\right)^2 \frac{c^2}{v_{\max}^2 - \frac{2GM_\odot}{r_{\min}}}}. \qquad (10.21)$$

Inserting the numerical values gives

$$dr_{\max}; = \frac{1}{\frac{(5\cdot 10^4)^2}{6.67\cdot 10^{-11}\cdot 2\cdot 10^{30}} - \frac{2}{0.5\cdot 1.5\cdot 10^{11}}} \pm \sqrt{\cdots}$$
$$= \frac{1}{1.87 \cdot 10^{-11} - 2.67 \cdot 10^{-11}} \pm \sqrt{\cdots} \qquad (10.22)$$
$$= 1.25 \cdot 10^{12} \pm \sqrt{1.56 \cdot 10^{24} - 1.4 \cdot 10^{22}}$$
$$= (1.25 \pm 1.25) \cdot 10^{12} m.$$

Here only the+sign is valid, because otherwise $r_{\max} < r_{\min}$.

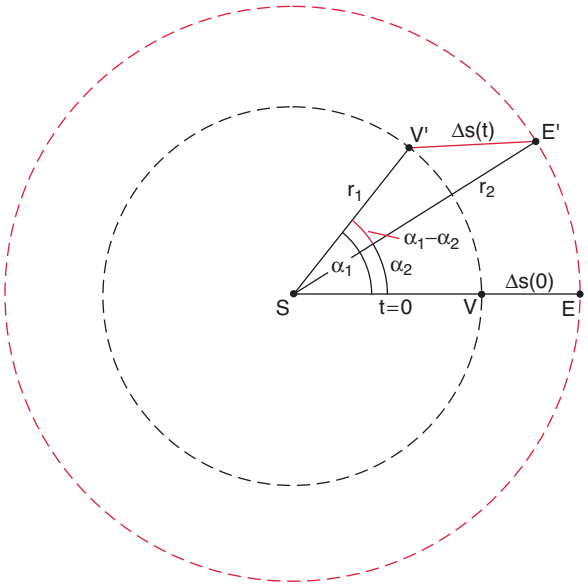


Fig. 10.3 To solution 3.2

$$\Rightarrow r_{\max} = 3 \cdot 10^{12}\text{m} = 20\,\text{AU}.$$

(b) The major semi-axis of the elliptical orbit is

$$+a = \frac{1}{2}(r_{\max} + r_{\min}) = \frac{1}{2}(20 + 0.5)\text{AU}.$$

The revolution period is then according to Kepler's 3. Law

$$T = \left(\frac{a_{\text{comet}}}{a_{\text{earth}}}\right)^{3/2}\text{years} = \left(\frac{10.25}{1}\right)^{3/2}\text{years} = 32.82\,\text{years}. \tag{10.23}$$

4.2 Division of (4.2c) by m and multiplying vectorial with $\boldsymbol{L} = m(\boldsymbol{r} \times \dot{\boldsymbol{r}})$ gives:

$$\boldsymbol{L} \times \ddot{\boldsymbol{r}} = m(\boldsymbol{r} \times \dot{\boldsymbol{r}}) \times \ddot{\boldsymbol{r}} = -G \cdot m\left(\frac{m + M_\odot}{r^3}\right)(\boldsymbol{r} \times \dot{\boldsymbol{r}}) \times \boldsymbol{r}. \tag{10.24}$$

The double vector product on the right side can be written as (see Vol. 1 Sects. 13–1.5.)

$$(\boldsymbol{r} \times \dot{\boldsymbol{r}}) \times \boldsymbol{r} = r^2\dot{\boldsymbol{r}} - (\boldsymbol{r} \cdot \dot{\boldsymbol{r}}) \cdot \boldsymbol{r}. \tag{10.25}$$

One therefore gets

$$\left(\boldsymbol{L} \times \ddot{\boldsymbol{r}}\right) = -G(m + M_\odot) \cdot m\left(\frac{\dot{\boldsymbol{r}}}{r} - \frac{\dot{r}}{r^2}\boldsymbol{r}\right). \tag{10.26}$$

The last bracket is equal to the time-derivative of **r**/r. Therefore, the integration gives with **L** = const.

$$(\boldsymbol{L} \times \dot{\boldsymbol{r}}) = -G \cdot m(m + M_\odot)\frac{\boldsymbol{r}}{r} + \boldsymbol{C}, \tag{10.27}$$

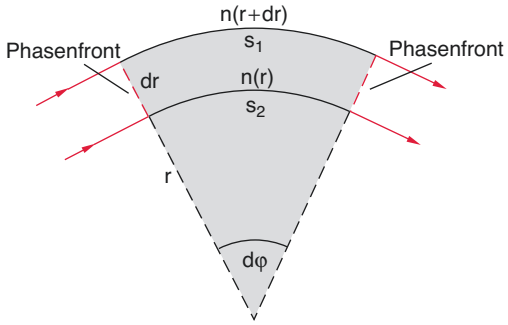


Fig. 10.4 To solution 3.3

where the constant vector **C** is the integration constant. Multiplication of (10.27) scalar with **r** gives for the left side

$$(\boldsymbol{L} \times \dot{\boldsymbol{r}}) \cdot \boldsymbol{r} = \boldsymbol{L} \cdot (\dot{\boldsymbol{r}} \times \boldsymbol{r}) = -L^2/m, \tag{10.28}$$

where $\boldsymbol{L} = m(\boldsymbol{r} \times \dot{\boldsymbol{r}})$.

This gives

$$L^2 = G\,m^2(m + M_\odot)r + C \cdot r \cdot \cos\varphi, \tag{10.29}$$

where φ is the angle between radius vector **r** and the x-axis.

Solving this equation for r yields

$$r = \frac{L^2}{Gm^2(m + M_\odot) + C\cos\varphi} = \frac{L^2/Gm^2(m + M_\odot)}{1 + C/Gm^2(m + M_\odot)\cos\varphi}. \tag{10.30}$$

The comparison with the polar equation with the ellipse (4.2d) gives

$$C = G\,m^2 \cdot \varepsilon(m + M_\odot). \tag{10.31}$$

For $\cos\varphi = -1$ the radius r becomes according to (4.2d) $r = a(1 + \varepsilon) \cdot (1 + m/M_\odot)$. Inserting this relation into (10.29) one obtains

$$L^2 = G\frac{m}{M_0}(m + M_\odot)^2\left(1 - \varepsilon^2\right)a. \tag{10.32}$$

The comparison with the squared Eq. (4.11) gives

$$\begin{array}{ll} \frac{\pi^2 a^4\left(1-\varepsilon^2\right)\cdot 4m^2}{T^2} & = G\frac{m^2}{M_0}(m + M_\odot)^2\left(1 - \varepsilon^1\right)a \\ \Rightarrow \frac{a^3}{T^2} & = \frac{G(m+M_\odot)^2}{4M_\odot\cdot\pi^2} \\ \frac{T^2}{a^3} & = \frac{4\pi^2 M_\odot}{G(m+M_\odot)^2}. \end{array} \tag{10.33}$$

4.3 The apparent visual brightness of a planet is determined by its distance from the sun and from the earth. It receives from the sun the radiation power

$$\frac{\text{d}W_1}{\text{d}t} = L_\odot \cdot \frac{\pi R_\text{P}^2}{\pi r_\text{P}^2} \tag{10.34}$$

with R_P = radius of the planet, r_P = distance sun-planet. With the albedo A the power

$$\frac{dW_2}{dt} = L_\odot \cdot \frac{R_P^2}{r_P^2} \cdot A \tag{10.35}$$

is diffusely backscattered by the planet. The earth receives then the power flux density

$$\Phi = \frac{1}{4\pi r_{EP}^2} \cdot \frac{dW_2}{dt}. \tag{10.36}$$

from the planet. The apparent magnitude of the planet is then

$$\begin{aligned} &m = -2.5 \cdot \log(\Phi/\Phi_0). \\ &\Rightarrow \text{for } r_1 = 1.55\text{AU} \Rightarrow r_{EP} = 0.55\text{AU} \\ &\Rightarrow \Phi_1 = \frac{1}{4\pi\,(0.55\,\text{AU})^2} \cdot L_\odot \cdot \frac{R_P}{(1.55\text{AU})^2} \cdot A \end{aligned} \tag{10.37}$$

For $r_2 = 1.64\,\text{AU}$

$$\Rightarrow \Phi_2 = \frac{1}{4\pi\,(0.64\,\text{AU})^2} \cdot L_\odot \cdot \frac{R_P}{(1.64\,\text{AU})^2} \cdot A. \tag{10.38}$$

$$\begin{aligned} m_1 - m_2 &= -2.5 \cdot \log \Phi_1/\Phi_2 \\ &= -2.5 \cdot \log \tfrac{1.64^2 \cdot 0.64^2}{1.55^2 \cdot 0.55^2} \\ &= -2.5 \cdot \log 1.516 = -0.45 \\ \Rightarrow m_2 &= m_1 + 0.45 = -1.6^m + 0.45^m = 1.15^m. \end{aligned} \tag{10.39}$$

4.4 According to the solution of (4.3) the radiation flux received from the planetoid is

$$\Phi_P = \frac{1}{4\pi r_{EP}^2} \cdot L_\odot \cdot A \cdot \frac{R_P^2}{r_P^2}. \tag{10.40}$$

The radiation flux, received by the earth from the sun is

$$\begin{aligned} \Phi_\odot &= \tfrac{1}{4\pi r_E^2} \cdot L_\odot \\ \Rightarrow \tfrac{\Phi_P}{\Phi_\odot} &= \tfrac{r_E^2 \cdot A \cdot R_P^2}{r_{EP}^2 r_P^2} = \tfrac{(1\,\text{AU})^2 \cdot A \cdot R_P^2}{(2\,\text{AU})^2 \cdot (3\,\text{AU})^2} \\ &= \tfrac{A \cdot R_P^2}{36 \cdot (1\,\text{AU})^2} \\ \Rightarrow R_P^2 &= \tfrac{36}{A} \tfrac{\Phi_P}{\Phi_\odot} \cdot (1\,\text{AU})^2 \\ m_P - m_\odot &= -2.5 \cdot \log \Phi_P/\Phi_\odot. \end{aligned} \tag{10.41}$$

The apparent magnitude of the sun can be determined from its luminosity L and its distance from the earth. It is $m_\odot = -26.8^m$. Therefore

$$\begin{aligned} m_P - m_\odot &= +10.0^m + 26.8^m = +36.8^m \\ \Rightarrow \Phi_P/\Phi_\odot &= 10^{-0.4 \cdot 38.6} = 10^{-14.72} \\ \Rightarrow R_P &= 6/\sqrt{A} \cdot 10^{-7.36}\text{AE} \\ &= 6/\sqrt{0.15} \cdot 4.4 \cdot 10^{-8} \cdot 1.5 \cdot 10^{11}\text{m} \\ &= 1.0 \cdot 10^5\text{m} = 100\,\text{km}. \end{aligned} \tag{10.42}$$

For $\Delta A = 30\% \Rightarrow \Delta R = 15\%$.

4.5 .(a) If R_0 is the radius without rotation, the mass M is

$$M = \frac{4}{3}\pi \bar{\varrho} R_0^3,$$

where $\bar{\varrho}$ is the mean density. Due to the centrifugal force of the rotating earth the radius at the equator is widened to

$$R_0 \to R_0 + \Delta R = a.$$

For the equilibrium is the difference of the gravitational forces without and with widening equal to the centrifugal force. This gives the relation

$$\begin{aligned} &G \cdot M \cdot m \left(\frac{1}{R_0^2} - \frac{1}{(R_0 + \Delta R)^2} \right) = m\omega^2 (R_0 + \Delta R) \\ &\Rightarrow \frac{4}{3}\pi \bar{\varrho} R_0^3 G \left[\frac{2R_0 \Delta R + \Delta R^2}{R_0^2} \right] = \omega^2 (R_0 + \Delta R)^3 \\ &\Rightarrow \frac{4}{3}\pi \bar{\varrho} R_0^3 G \left[2R_0^2 \Delta R + R_0 \Delta R^2 \right] \\ &\quad \approx \omega^2 \left[R_0^3 + 3R_0^2 \Delta R + \cdots \right] \\ &\quad \approx \frac{4}{3}\pi \bar{\varrho} G \cdot 2R_0^2 \Delta R \approx \omega^2 \left[R_0^3 + 3R_0^2 \Delta R \right] \end{aligned} \tag{10.43}$$

$$\Rightarrow \Delta R \approx \frac{\omega^2 R_0}{\frac{8}{3}\pi \bar{\varrho} G - 3\omega^2}. \tag{10.44}$$

Inserting the numerical value for the rotation period

$$\begin{aligned} T &= 10\text{h}\,30\,\text{min} \\ \Rightarrow \omega &= \tfrac{2\pi}{37.800}\text{s}^{-1} = 1.662 \cdot 10^{-4}\text{s}^{-1}, \\ \bar{\varrho} &= 1.3 \cdot 10^3\,\text{kg/m}^3, \\ R_0 &= \left(\tfrac{3M}{4\pi \varrho_0} \right)^{1/3} = 7.1 \cdot 10^7\text{m} \\ &\Rightarrow \Delta R \\ &= \tfrac{1.662^2 \cdot 10^{-8} \cdot 7.1 \cdot 10^7}{\frac{8}{3}\pi \cdot 1.3 \cdot 10^3 \cdot 6.67 \cdot 10^{-11} - 3 \cdot 1.662^2 \cdot 10^{-8}} \\ &= \tfrac{1.96}{7.26 \cdot 10^{-7} - 0.83 \cdot 10^{-7}} \\ &= 3.0 \cdot 10^6\text{m} = 0.046 R_0. \end{aligned} \tag{10.45}$$

The widening of the radius due to the rotation of a homogeneous gas sphere amounts to 4.3%.

Because of the conservation of the total mass, we get

$$\begin{aligned} M = \tfrac{4}{3}\pi \bar{\varrho} R_0^3 &= \tfrac{4}{3}\pi \bar{\varrho} a^2 \cdot b \,\text{mit}\, a = R_0 + \Delta R \\ \Rightarrow b &= \tfrac{\frac{4}{3}\pi \bar{\varrho} R_0^3}{\frac{4}{3}\pi \bar{\varrho} a^2} = \tfrac{R_0^3}{(R_0 + \Delta R)^2} \\ &\approx R_0 \cdot \tfrac{1}{\left(1 + \frac{\Delta R}{R_0}\right)^2} \approx R_0 \left(1 - \tfrac{2\Delta R}{R_0}\right) \\ &\approx 0.914 R_0. \end{aligned} \tag{10.46}$$

The oblateness is then

$$\begin{aligned} \left(R_{\text{equator}} - R_{\text{pol}}\right)/R_{\text{equator}} &= \tfrac{a-b}{a} = \tfrac{1.046 - 0.914}{1.046} \\ &= 0.126. \end{aligned} \tag{10.47}$$

This shows that Jupiter is not a homogeneous gas sphere because its obleteness is smaller.

(b) If the density

$$\varrho = \varrho_0[1 - r/R]$$

decreases linear from the center to the edge the mass relation $M = \frac{4}{3}\pi\overline{\varrho}_0 R^3$ must be replaced by

$$\begin{aligned} M &= 4\pi\varrho_0 \int_0^R (1 - r/R) r^2 \mathrm{d}r \\ &= \tfrac{4\pi}{12}\varrho_0 R^3 = \tfrac{4}{3}\pi\overline{\varrho} R^3 \\ \Rightarrow \overline{\varrho} &= \tfrac{1}{4}\varrho_0 \end{aligned} \tag{10.48}$$

where ϱ_0 is the density in the center $r = 0$.

Analog to (10.44) one obtains (10.43) but instead of (10.43a) one now obtains

$$\begin{aligned} \tfrac{4\pi}{12}\varrho_0 \cdot G \cdot 2R_0^2 \Delta R &\approx \omega^2 \left[R_0^3 + 3R_0^2 \Delta R + \cdots\right] \\ \Rightarrow \Delta R &= \tfrac{\omega^2 R_0}{\frac{2\pi}{3}\varrho_0 G - 3\omega^2}. \end{aligned} \tag{10.49}$$

With $\varrho_0 = 4\overline{\varrho}$ the same value for ΔR is obtained as in (10.44).

(c) The mass of Jupiter is for instance, obtained from the revolution period T and the distance r_{Ga} of the moon Ganymede.

$$\begin{aligned} G\tfrac{m \cdot M}{r_{Ga}^2} &= m \cdot \omega^2 \cdot r_{Ga} \\ \Rightarrow M &= \omega^2 \cdot \tfrac{r_{Ga}^3}{G}. \\ T &= 7.1546\,\text{Tage} \\ \Rightarrow \omega &= \tfrac{2\pi}{7.1546 \cdot 24 \cdot 3600} \\ &= 1.016 \cdot 10^{-5}\text{s}^{-1} \\ \Rightarrow M &= \tfrac{1.016^2 \cdot 10^{-10} \cdot 1.0705^3 \cdot 10^{27}}{6.67 \cdot 10^{-11}}\text{kg} \\ &= 1.90 \cdot 10^{27}\text{kg}. \end{aligned} \tag{10.50}$$

4.6 The small body with mass M_2 and radius R should rotate with the angular velocity ω around the center O_1 of M_1. The total force onto its center O_2 is zero, because the gravitational force i is equal to the centripetal force.

$$\begin{aligned} G \cdot \tfrac{M_1 M_2}{r^2} &= M_2 \omega^2 \cdot r \\ \Rightarrow \omega^2 &= G \cdot \tfrac{M_1}{r^3}. \end{aligned} \tag{10.51}$$

The Roche limit is reached, if the difference $\Delta\boldsymbol{F} = \boldsymbol{F}_1 - \boldsymbol{F}_2$ of the forces

$$\begin{aligned} \boldsymbol{F}_1 &= +\left[G \cdot \tfrac{m \cdot M_1}{(r-R)^2} - m\omega^2(r - R)\right]\hat{\boldsymbol{r}}, \\ \boldsymbol{F}_2 &= +\left[G \cdot \tfrac{m \cdot M_1}{(r+R)^2} - m\omega^2(r + R)\right]\hat{\boldsymbol{r}} \end{aligned} \tag{10.52}$$

acting on the probe masses m in the points P_1, P_2 equals the gravitational force

$$\Delta F = F_3 = G \cdot \frac{2m \cdot M_2}{R^2} \tag{10.53}$$

for $r < r_s$ is $|\Delta F| > |F_3|$. This means the body M_2 is pulled apart by the difference $\boldsymbol{F}_1 - \boldsymbol{F}_2$. The condition for the limit $r = r_s$ is therefore

$$\begin{aligned} G \cdot M_1 \left[\tfrac{1}{(r_s - R)^2} - \tfrac{1}{(r_s + R)^2} + \tfrac{2\omega^2 R}{GM_1}\right] \\ = \tfrac{2G \cdot M_2}{R^2}. \end{aligned} \tag{10.54}$$

With $\omega^2 = G \cdot M_1 / r_s^3$ it follows:

$$\begin{aligned} \tfrac{1}{(r_s - R)^2} - \tfrac{1}{(r_s + R)^2} + \tfrac{2R}{r_s^3} &= \tfrac{2M_2}{M_1 R^2} \\ \Rightarrow \tfrac{4r_s R}{\left(r_s^2 - R^2\right)^2} + \tfrac{2R}{r_s^3} &= \tfrac{2M_2}{M_1 R^2}. \end{aligned} \tag{10.55}$$

For $R \ll r_s$ this can be approximated by

$$\frac{2R}{r_s^3} + \frac{R}{r_s^3} = \frac{M_2}{M_1 R^2} \Rightarrow r_s = R \cdot \sqrt[3]{M_1/M_2}. \tag{10.56}$$

With the mass density ϱ_2 of the small body with mass $M_2 = \frac{4}{3}\pi\varrho R^3$ we get

$$r_s = \sqrt[3]{\frac{3M_1}{4\pi\varrho_2}}. \tag{10.57}$$

The Roche limit therefore depends on the density ϱ_2 of the small and the mass M_1 of the large body.

In the case of the Jupiter moon Io is $\varrho_2 = 3.5 \cdot 10^3\text{kg/m}^3$. Inserting the Jupiter mass $M_1 = 1.9 \cdot 10^{27}\text{kg}$ one obtains for the Roche limit

$$\begin{aligned} \Rightarrow r_s &= \sqrt[3]{\tfrac{5.7 \cdot 10^{27}}{4\pi \cdot 3.5 \cdot 10^3}}\text{m} = 0.5 \cdot 10^8\text{m} \\ &= 5 \cdot 10^4\text{km}. \end{aligned} \tag{10.58}$$

The radius of the orbit of Io is $a = 42.2 \cdot 10^4\text{km}$. Io is therfore outside the Roche limit (otherwise it would no longer exist).

4.7 We regard three masses m_1, m_2, m_3 at the corners of an equilateral triangle with side-length r. In case of the Trojans is m_1 (sun) and m_2 (Jupiter), very large to m_3 (Trojans). Relative to the common center-of-mass S, which nearly lies on the line m_1-m_2 and divides it with the ratio m_1/m_2. The Newton's equation of motion is

$$m_i \cdot \ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i. \tag{10.59}$$

The whole system rotates with the angular velocity $\boldsymbol{\omega}$ around S. All three masses move in a plane perpendicular to $\boldsymbol{\omega}$.

The side-lengths r of the triangle can be time dependent, i.e. the triangle "breathes". Because the angular momentum must be preserved, also the angular velocity must be time dependent, When the side-lenghts increase, ω decreases, but it does not change its direction. When we write (10.59) in the rotating coordinate system, Coriolis- and centrifugal

forces must be taken into account (see Vol. 1 Sect. 3.3) and (10.59) becomes

$$\frac{\partial^2 \boldsymbol{r}_i}{\partial t^2} + 2\left(\boldsymbol{\omega} \times \frac{\partial \boldsymbol{r}_i}{\partial t}\right) + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}_i) + (\dot{\boldsymbol{\omega}} \times \boldsymbol{r}_i) = \boldsymbol{F}_i/m_i. \tag{10.60}$$

If the masses always stay at the corners of the triangle their radius vectors are

$$\boldsymbol{r}_i = \lambda(t) \cdot r_{i0},$$

where $\lambda(t)$ is a scalar time dependent parameter. Inserting this into (10.60) we obtain with $\dot{\boldsymbol{r}}_i = \dot{\lambda} \cdot \boldsymbol{r}_{i0}, \ddot{r}_i = \ddot{\lambda} \cdot \boldsymbol{r}_{i0}$ after the vectorial multiplication of (10.60) with $\mathbf{r_i}$

$$\begin{aligned}&\lambda \cdot \ddot{\lambda}\,(\boldsymbol{r}_{i0} \times \boldsymbol{r}_{i0}) \\ &+ 2\lambda\dot{\lambda}\boldsymbol{r}_{i0} \times (\boldsymbol{\omega} \times \boldsymbol{r}_{i0}) + \lambda^2 \boldsymbol{r}_{i0} \times (\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}_{i0})) \\ &+ \lambda^2 \boldsymbol{r}_{i0} \times (\dot{\boldsymbol{\omega}} \times \boldsymbol{r}_{i0}) = (\lambda/m_i)(\boldsymbol{r}_{i0} \times \boldsymbol{F}_i).\end{aligned} \tag{10.61}$$

because $\boldsymbol{\omega} \perp \boldsymbol{r}_{i0}$ we get

$$\begin{aligned}&\Rightarrow\ \boldsymbol{r}_{i0} \times (\boldsymbol{\omega} \times \boldsymbol{r}_{i0}) = r_{i0}^2 \boldsymbol{\omega} - (\boldsymbol{r}_{i0} \cdot \boldsymbol{\omega})\boldsymbol{r}_{i0} = r_{i0}^2 \cdot \boldsymbol{\omega} \\ &\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}_{i0}) = -\omega^2 \cdot \boldsymbol{r}_{i0} \Rightarrow \boldsymbol{r}_{i0} \times \omega^2 \boldsymbol{r}_{i0} = 0.\end{aligned} \tag{10.62}$$

Furthermore is $\dot{\boldsymbol{\omega}} \parallel \boldsymbol{\omega}$. Therefore, many terms in (10.61) become zero and it remains

$$\begin{aligned}&2\lambda\dot{\lambda} r_{i0}^2 \boldsymbol{\omega} + \lambda^2 r_{i0}^2 \dot{\boldsymbol{\omega}} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\lambda^2 r_{i0}^2 \boldsymbol{\omega}\right] \\ &= (\lambda/m_i)(\boldsymbol{r}_{i0} \times \boldsymbol{F}_i).\end{aligned} \tag{10.63}$$

Summing (10.63) over the three indices i, the sum over the right side becomes zero, because the angular momentum is zero in the corotating system.

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left[\lambda^2 \cdot \boldsymbol{\omega} \cdot \sum_{i=1}^{3} m_i r_{i0}^2\right] = 0. \tag{10.64}$$

because

$$\sum m_i r_{i0}^2 = const. \tag{10.65}$$

Then it follows from (10.63) that for each individual force it must

$$\begin{aligned}&\boldsymbol{F}_i \boldsymbol{r}_{i0} \times \boldsymbol{F}_i = 0 \\ &\Rightarrow \boldsymbol{r}_{i0} \parallel \boldsymbol{F}_i.\end{aligned}$$

The resultant force points for each of the three masses towards the center of mass S. Therefore, the force acting on the mass m_i is

$$\begin{aligned}\boldsymbol{F}_i &= -G \cdot m_i \cdot M \frac{\boldsymbol{r}_i}{r^3} \\ &= -G \cdot m_1 \cdot M \left(\frac{r_{i0}}{r_0}\right)^3 \frac{\boldsymbol{r}_i}{r_i^3},\end{aligned} \tag{10.66}$$

where $M = \sum m_i$ and $r = \lambda r_0$ is the side-length of the triangle. Each of the three masses moves around the center of mass induced by the centrifugal force. It must move on a circle or an ellipse.

It is possible to calculate the potential of this constraint three body problem, although this demands some efforts (see for example K. Stumpf: Himmelsmechanik VEB Verlag der Wissenschften Berlin 1965).

Such calculations show, that the points L_4 and L_5 of the equilateral triangles SJL_4 and SJL_5 are minima of the potential energy (Libration points) while in the points L_1, L_2 and L_3 saddle points of the potential energy occur.

Small masses m therefore can perform stable orbits (circles or ellipses) around the libration points L_4 and L_5. In the points L_1, L_2, L_3 they experience a metastable equilibrium. Elongations away from these points result in unstable orbits.

The reason for the minima of E_{pot} are the superposition of gravitational forces, centrifugal forces and Coriolis forces acting at the synchronous rotation of m_2 and m_3 around the sun.

4.8 When a space probe passes a hypothetical fixed planet (which would not orbit around the sun) it would not gain kinetic energy, because the gain, achieved during the approach would be lost again after the approach. Its total energy would not change, but its flight time would be shortened because its velocity would be larger during the approach.

The real planet has its own velocity

$$v_P = \omega \cdot r_P = \frac{2\pi r_P}{T} \propto \frac{1}{\sqrt{r_P}}, \tag{10.67}$$

which is proportional to $1/\sqrt{r_P}$, where r_p is ist distance from the sun.

If v_s is the velocity of the space probe in the coordinate system with the sun in the center. In a system with its origin in the center of the planet the velocity of the space probe is

$$\boldsymbol{v}_r = \boldsymbol{v}_S - \boldsymbol{v}_P,$$

and therefore its kinetic energy, before entering the attraction range of the planet is

$$E_{kin}^{rel} = \frac{1}{2} m_S \boldsymbol{v}_r^2 = \frac{1}{2} m_S \left(v_S^2 + v_P^2 - 2\boldsymbol{v}_S \cdot \boldsymbol{v}_P\right). \tag{10.68}$$

In the comoving system the energy of the space probe at the exit of the attraction range must be the same as for the entrance.

$$\begin{aligned}&\Rightarrow \tfrac{1}{2} m_S |\boldsymbol{v}_S - \boldsymbol{v}_P|^2 = \tfrac{1}{2} m_S |\boldsymbol{v}'_S - \boldsymbol{v}_P^2| \\ &\Rightarrow v_S^2 - 2\boldsymbol{v}_S \cdot \boldsymbol{v}_P = v_S'^2 - 2\boldsymbol{v}_S \cdot \boldsymbol{v}_P.\end{aligned} \tag{10.69}$$

This gives for the energy in the system with its origin in the sun:

$$E'_{\text{kin}} = \tfrac{1}{2}m_S v'^2_S = \tfrac{1}{2}m_S\left[v_S^2 + 2\boldsymbol{v}_P\left(\boldsymbol{v}'_S - \boldsymbol{v}_S\right)\right] = E_{\text{kin}} + m_S\boldsymbol{v}_P\cdot\left(\boldsymbol{v}'_S - \boldsymbol{v}_S\right). \tag{10.70}$$

If v_s had been antiparallel to v_p and v_s ' perpendicular to v_p then is $\boldsymbol{v}_P\cdot\boldsymbol{v}_S = -v_P\cdot v_S$ and $\boldsymbol{v}_P\cdot\boldsymbol{v}'_S = 0$. We obtain in this case

$$E'_{\text{kin}} = E_{\text{kin}} + m_S\cdot v_P\cdot v_S = E_{\text{kin}}\left(1 + 2\frac{v_P}{v_S}\right). \tag{10.71}$$

This technique, where a space probe wins energy by passing a planet is called "**swing by method**" (Fig. 10.5).

4.9 The potential energy of a spherical shell with thickness dr at the distance r from the center of a homogeneous spherical mass is

$$\mathrm{d}E_{\text{pot}} = \frac{-G\cdot M(r)\cdot 4\pi\varrho r^2\mathrm{d}r}{r}. \tag{10.72}$$

With $M(r) = \frac{4}{3}\pi\varrho r^3 M(r) = \frac{4}{3}\pi\varrho r^3$

$$\Rightarrow E_{\text{pot}} = -\tfrac{16}{3}\pi^2\varrho^2\cdot G\cdot\int_0^R r^4\mathrm{d}r = -\tfrac{16}{15}\pi^2\varrho^2 G\cdot R^5. \tag{10.73}$$

With

$$M = \tfrac{4}{3}\pi\varrho R^3 \Rightarrow M^2 = \tfrac{16}{9}\pi^2\varrho^2 R^6 \Rightarrow E_{\text{pot}} = -\tfrac{3}{5}G\cdot\tfrac{M^2}{R}. \tag{10.74}$$

4.10 We regard as small cylinder with ground area dA and length dr. The radiation passes through this cylinder and is partly absorbed. When the incident spectral intensity I_ν perpendicular to the ground area dA is attenuated by $\mathrm{d}I_n$ along the path dr due to absorption, the energy in the cylinder changes per sec and solid angle dΩ by the amount

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \mathrm{d}I_\nu\cdot\mathrm{d}A\cdot\mathrm{d}\nu\cdot\mathrm{d}\Omega. \tag{10.75}$$

on the other side the absorbed radiation power is for an absorption coefficient $\alpha(\nu)$

$$\left(\frac{\mathrm{d}W}{\mathrm{d}t}\right)_{\text{abs}} = \alpha_\nu I_\nu\mathrm{d}r\cdot\mathrm{d}A\cdot\mathrm{d}\nu\cdot\mathrm{d}\Omega \tag{10.76}$$

The power, emitted by the volume element dAdr is

$$\left(\frac{\mathrm{d}W}{\mathrm{d}t}\right)_{\text{em}} = \varepsilon_\nu\mathrm{d}\nu\cdot\mathrm{d}r\cdot\mathrm{d}A\cdot\mathrm{d}\Omega, \tag{10.77}$$

where ε_ν is the spectral emission coefficient? The difference of (10.76) and (10.77) must be equal to (10.75).

$$\Rightarrow \frac{\mathrm{d}W}{\mathrm{d}t} = -\left(\frac{\mathrm{d}W}{\mathrm{d}t}\right)_{\text{abs}} + \left(\frac{\mathrm{d}W}{\mathrm{d}t}\right)_{\text{em}} \tag{10.78}$$
$$\Rightarrow \mathrm{d}I_\nu = -\alpha_\nu I_\nu\mathrm{d}r + \varepsilon\mathrm{d}r.$$

The spectral radiation density of a black body (see Vol. 2. Section 12.2)

$$S_\nu = \varepsilon_\nu/\alpha_\nu, \tag{10.79}$$

Is equal to the ratio of emission- to absorption coefficient. From (10.78) we then obtain

$$\frac{1}{\alpha_\nu}\frac{\mathrm{d}I_\nu}{\mathrm{d}r} = -I_\nu + S_\nu. \tag{10.80}$$

This is the basic equation for the radiation transport.

If the radiation propagates in the direction ϑ against the normal to the ground area we obtain from (10.80)

$$\cos\vartheta\frac{\mathrm{d}I_\nu}{\mathrm{d}r} = -\alpha_\nu I_\nu + \alpha_\nu S_\nu. \tag{10.81}$$

Multiplication with $\cos\vartheta$ and integration over all directions ϑ gives:

$$\int_0^{2\pi}\cos^2\vartheta\frac{\mathrm{d}I_\nu}{\mathrm{d}r}\mathrm{d}\vartheta = -\alpha_\nu\int\cos\vartheta I_\nu\mathrm{d}\vartheta + \alpha_\nu\int\cos\vartheta S_\nu\mathrm{d}\vartheta. \tag{10.82}$$

Since S_ν is independent of ϑ (the radiation density of a black body is isotrop!) the last term in (10.82) becomes zero. On the left side of (10.82)

$$I_\nu = \frac{c}{4\pi}w_\nu \tag{10.83}$$

can be replaced by the energy density w_ν (see Vol. 2, Sect. 12.4). We then get the Planck formula for the radiation density

$$I_\nu = \frac{2h\nu^3}{c^2}\frac{1}{\mathrm{e}^{h\nu/kT} - 1}. \tag{10.84}$$

Integration over all frequencies yields the Stefan-Boltzmann law

$$\int_\vartheta\int_\nu\cos^2\vartheta I_\nu\mathrm{d}\vartheta\mathrm{d}\nu = 2\pi\sigma T^4. \tag{10.85}$$

Integration of (10.82) over all frequencies ν gives

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(T^4\right) = -\frac{1}{2\pi\sigma}\int_\nu\int_\vartheta\alpha_\nu I_\nu\cos\vartheta\,\mathrm{d}\nu\mathrm{d}\vartheta. \tag{10.86}$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(T^4\right) = \frac{\mathrm{d}}{\mathrm{d}T}\left(T^4\right)\cdot\frac{\mathrm{d}T}{\mathrm{d}r} = 4T^3\frac{\mathrm{d}T}{\mathrm{d}r} \tag{10.87}$$

we get

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{\alpha_\nu}{8\pi\sigma T^3} \int_\nu \int_\vartheta \alpha_\nu I_\nu \mathrm{d}\nu \cos\vartheta \,\mathrm{d}\vartheta\,. \tag{10.88}$$

The integral equals the radiation flux $\Phi(r)$through the area $4\pi r^2$. Introducing the luminosity $L(r) = 4\pi r^2 \cdot \Phi(r)$ this gives with the absorption coefficient per mass unit $\kappa = \frac{1}{\varrho}\int \alpha_\nu \mathrm{d}\nu$ the relation

$$\frac{\mathrm{d}T}{\mathrm{d}r} = C \cdot \frac{L(r)}{2\pi r^2}\frac{1}{T^3} \cdot \kappa \cdot \varrho. \tag{10.89}$$

4.11 A body with mass m experiences at the sun surface the gravitational force

$$F_\mathrm{G} = -G \cdot \frac{m \cdot M_\odot}{R_\odot^2}. \tag{10.90}$$

This corresponds to the acceleration

$$a = F_\mathrm{G}/m = -\frac{G \cdot M_\odot}{R_\odot^2} = 274\,\mathrm{m/s^2} \tag{10.91}$$

(28 times the acceleration on the earth surface).

On a layer with thickness $\mathrm{d}r$, density $\varrho(R_\odot) = \varrho_0$ and area A the force

$$F_\mathrm{G} = -\frac{G \cdot M_\odot}{R_\odot^2} \cdot \varrho_0 \cdot A \cdot \mathrm{d}r. \tag{10.92}$$

is acting. The pressure onto a sphere with radius $r = R_\odot$ is

$$p(R_\odot) = -\frac{G \cdot M_\odot}{R_\odot^2} \int_{R_\odot}^{\infty} \varrho \mathrm{d}r, \tag{10.93}$$

which is caused by the layers above the sun edge at $R = R_\odot$ (chromosphere and Corona). Since the density of these layers is small, the pressure at the outer edge of the photosphere at $R = R_\odot$ is mall and can be neglected compared with the pressure in the center of the sun. Its value can be obtained as follows:

The density ϱ can be written as

$$\begin{aligned} \varrho(r) &= \varrho(R_\odot) \cdot \mathrm{e}^{-(r-R_\odot)/a} \\ \Rightarrow \varrho_0 \int_{R_\odot}^{\infty} \mathrm{e}^{-(r-R_\odot)/a} &= a \\ \Rightarrow p(R_\odot) &= a \cdot \varrho(R_\odot) \cdot \tfrac{GM_\odot}{R_\odot^2}. \end{aligned} \tag{10.94}$$

Inserting the numerical values $a = 10^5$ m and $\varrho(R_\odot) = 3 \cdot 10^{-5}\,\mathrm{kg/m^3}$ (Fig. 4.112) yields

$$\begin{aligned} p(R_\odot) &= 10^5 \cdot 3 \cdot 10^{-5} \cdot \tfrac{6.67 \cdot 10^{-11} \cdot 2 \cdot 10^{30}}{7 \cdot 2 \cdot 16} \tfrac{\mathrm{N}}{\mathrm{m^2}} \\ &= 8.2 \cdot 10^2 \mathrm{N/m^2}. \end{aligned} \tag{10.95}$$

4.12 The escape velocity $\mathrm{v_e}$ of a body with mass m from a celestial object with mass M and radius R is given by

$$\begin{aligned} &\mathrm{m/2ve^2} \geq \mathrm{G \cdot m \cdot M/R} \\ &\rightarrow \mathrm{V_e} \geq \mathrm{(2GM/R)^{1/2}} \end{aligned}$$

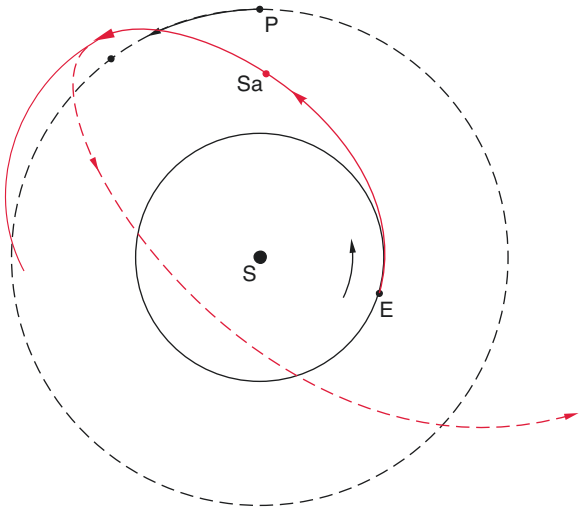


Fig. 10.5 To solution 4.8

(a) Sun: M=
(b) Jupiter: M=
(c) Earthmoon: M=4.13 The potential energy of a spherical shell with the mass $\mathrm{d}m = 4\pi\varrho r^2 \mathrm{d}r$ in a star with radius R is

$$\mathrm{d}E_\mathrm{pot} = -\frac{G \cdot M(R) \cdot 4\pi\varrho r^2 \mathrm{d}r}{r}, \tag{10.96}$$

since the layers with $r' > r$ do not contribute to the potential energy. The total potential energy is then

$$\begin{aligned} E_\mathrm{pot} &= -G \cdot 4\pi \int_0^R M(R)\varrho \cdot r \mathrm{d}r \\ &= -4\pi G \int_0^R \tfrac{4\pi}{3}\varrho r^3 \cdot \varrho r \mathrm{d}r \\ &= -\tfrac{16\pi^2}{3} G \cdot \int_0^R \varrho^2 r^4 \mathrm{d}r. \end{aligned} \tag{10.97}$$

For a homogeneous star (ϱ=const.) is

$$\begin{aligned} E_\mathrm{pot} &= -\tfrac{16\pi^2}{3} G \cdot \varrho^2 \cdot \tfrac{1}{5} R^5 \\ &= -\tfrac{3G}{5}\tfrac{M^2(R)}{R}. \end{aligned} \tag{10.98}$$

while for a density distribution

$$\begin{aligned} \varrho(r) &= \varrho_0 \cdot \left(1 - \tfrac{r}{R}\right)^2 \\ \Rightarrow M(r) &= 4\pi\varrho_0 \int_0^r \left(1 - \tfrac{r'}{R}\right)^2 r'^2 \mathrm{d}r' \\ &= \tfrac{2\pi}{15}\varrho_0 r^3, \end{aligned} \tag{10.99}$$

This gives for the potential energy

$$\begin{aligned} E_\mathrm{pot} &= -G \cdot 4\pi \int_0^R \tfrac{2\pi}{15}\varrho_0 r^3 \cdot \varrho_0 \left(1 - \tfrac{r}{R}\right)^2 r \mathrm{d}r \\ &= -\tfrac{2}{7} G \tfrac{M^2}{R}. \end{aligned} \tag{10.100}$$

The gravitational potential energy therefore depends only slightly on the density distribution.

During the contraction of a star from the radius R to the smaller radius R_0 the potential energy becomes for the density distribution $\varrho(r) = \varrho_0 \cdot \left(1 - \left(\frac{r}{R}\right)^2\right)$

$$E_{\text{pot}} = \left(\frac{2}{7}\right) G \cdot M^2 \left(\frac{1}{R_0} - \frac{1}{R}\right) \tag{10.101}$$

For the contraction of a gas cloud with $R \gg R_0$ the potential energy becomes

$$E_{\text{pot}} = \left(\frac{2}{7}\right) G \cdot M^2 / R_0$$

This can be also obtained by the following estimation:

If one adds the mass dm from a far distance to the mass $m(R_0)$ contained within the sphere with radius R_0 the potential energy is increased by

$$\begin{aligned} \mathrm{d}E_{\text{pot}} &= \mathrm{d}m\left[\Phi_{\text{pot}}(R) - \Phi_{\text{pot}}(R_0)\right] \\ &= G \cdot \mathrm{d}m \cdot m(R_0)\left(\tfrac{1}{R_0} - \tfrac{1}{R}\right) \\ &\approx \tfrac{G \cdot m(R_0) \cdot \mathrm{d}m}{R_0}. \end{aligned} \tag{10.102}$$

Increasing the mass from m = 0 to $M(R_0)$ the potential energy is

$$E_{\text{pot}} = \frac{G}{R_0} \int_0^M m \mathrm{d}m = \frac{1}{2} G \frac{M^2}{R_0}. \tag{10.103}$$

4.14.

Für die Abhängigkeit des Druckes p vom Radius r innerhalb einer homogenen Massenkugel mit Masse M und Radius R gilt:

$$\frac{\partial p}{\partial r} = -g \cdot \varrho \quad \text{mit} \quad g = G \cdot \frac{m(r)}{r^2} \varrho. \tag{10.104}$$

M i t $m(r) = \frac{4}{3}\pi\varrho r^3 \Rightarrow \frac{\partial m}{\partial r} = 4\pi\varrho r^2 m(r) = \frac{4}{3}\pi\varrho r^3 \Rightarrow \frac{\partial m}{\partial r} = 4\pi\varrho r^2$

$$\Rightarrow \frac{\partial p}{\partial m} = \frac{\partial p}{\partial r} \cdot \frac{\partial r}{\partial m} = \frac{G \cdot m \cdot \varrho}{r^2 \cdot 4\pi\varrho r^2} = -G \cdot \frac{m}{4\pi r^4}.$$

Ersetzen wir: $m(r) = \frac{4}{3}\pi\overline{\varrho} r^3 m(r) = \frac{4}{3}\pi\overline{\varrho} r^3$

$$\begin{aligned} \Rightarrow r^4 &= \left(m/\left(\tfrac{4}{3}\pi\overline{\varrho}\right)\right)^{4/3} \\ \Rightarrow \tfrac{\partial p}{\partial m} &= -\tfrac{G}{4\pi}\left(\tfrac{4}{3}\pi\overline{\varrho}\right)^{4/3} \cdot m^{-1/3} \\ \Rightarrow p_{\text{g}}(r=0) &= \tfrac{3}{2} \cdot \tfrac{G}{4\pi}\left(\tfrac{4}{3}\pi\overline{\varrho}\right)^{4/3} M^{2/3} \\ &= \tfrac{3}{8} G \cdot \pi^{1/3}\left(\tfrac{4}{3}\overline{\varrho}\right)^{4/3} \cdot \left(\tfrac{4}{3}\pi\overline{\varrho} R^3\right)^{2/3} \\ &= \tfrac{3}{8}\pi \cdot G \cdot \tfrac{16}{9}\overline{\varrho}^2 \cdot R^2 \\ p_{\text{g}} &= \tfrac{2}{3}\pi G \overline{\varrho}^2 \cdot R^2. \end{aligned}$$

Wenn dieser Gravitationsdruck durch die elastische Kompression des Materials im Erdinneren aufgefangen werden soll, dann muss

$$p_{\text{c}} \le \Delta p = K \cdot \Delta V / V \tag{10.105}$$

gelten, wobei K das Kompressionsmodul ist. Die Dichte im Zentrum ist etwa dreimal so groß wie die Dichte von Eisen-Nickel ohne Druck. Deshalb muss $\Delta V/V = 2/3$ sein.

$$\begin{aligned} \Rightarrow \tfrac{2}{3} &= \pi G \overline{\varrho}^2 R^2 \le \tfrac{2}{3} K \\ \Rightarrow R &\le \left(\tfrac{K}{\pi G \overline{\varrho}^2}\right)^{1/2} \\ &\le \left(\tfrac{3 \cdot 10^{11}}{\pi \cdot 6.67 \cdot 10^{-11} \cdot 16 \cdot 10^6}\right)^{1/2} \text{m} \\ &= 0.95 \cdot 10^7 \text{m} \approx 9500\,\text{km}. \end{aligned} \tag{10.106}$$

Der Radius der Erde ist aber nur 6378 km. In der Lösung ist nicht berücksichtigt worden, dass der Kompressionsmodul K vom Druck abhängt. Dies führt zu dem zu großen Wert von R.

Chapter 5

5.1 The luminosity of a star is

$$L = 4\pi R^2 \sigma \cdot T^4.$$

For a complete occultation of one component by the other in a binary system (the orbital plane of the two components must lie in the line of sight to the earth) the radiation power, received on earth is for the positions (Fig. 10.6)

(a) $L_1 \propto T_1^4$,
(b) $L_1 + L_2 \propto T_1^4 + T_2^4$,
(c) $L_2 \propto T_2^4$,
(d) $L_1 + L_2 \propto T_1^4 + T_2^4$.

For $T_1 > T_2$ the minimum is reached for position c).

5.2 If the pulsar is regarded as a spherical neutron star with homogeneous density ϱ = const., its rotation energy is

$$E_{\text{rot}} = \frac{1}{5} M R^2 \cdot \omega^2 = \frac{1}{5} \cdot \varrho \cdot \frac{4}{3}\pi R^5 \cdot \frac{4\pi^2}{\tau^2}. \tag{10.110}$$

Inserting the numerical values

$$M = 4 \cdot 10^{30}\text{kg}, R = 8 \cdot 10^3\text{m}, \tau = 1.2 \cdot 10^{-3}\text{s}$$
$$\Rightarrow \omega = 5.24 \cdot 10^3 \text{s}^{-1} \Rightarrow E_{\text{rot}} = 1.4 \cdot 10^{45}\text{J}.$$

The emitted radiation power is

$$\begin{aligned} P &= -\frac{\mathrm{d}E_{\text{rot}}}{\mathrm{d}t} = -2 \cdot \frac{1}{5} M R^2 \omega \cdot \frac{\mathrm{d}\omega}{\mathrm{d}t} \\ &= -2E_{\text{rot}} \cdot \frac{1}{\omega}\frac{\mathrm{d}\omega}{\mathrm{d}t} \\ \omega &= \tfrac{2\pi}{\tau} \Rightarrow \tfrac{\mathrm{d}\omega}{\mathrm{d}t} = -\tfrac{2\pi}{\tau^2}\tfrac{\mathrm{d}\tau}{\mathrm{d}t} \\ \tfrac{\mathrm{d}\tau}{\mathrm{d}t} &= \tfrac{10^{-5}\tau}{year} = \tfrac{10^{-5}\tau}{3.15 \cdot 10^7\text{s}} \\ \Rightarrow P &= 2E_{\text{rot}} \cdot \tfrac{10^{-5}}{3.15 \cdot 10^7\text{s}} \\ &= 1.4 \cdot 10^{45} \cdot 3.2 \cdot 10^{-13} \\ &= 4.5 \cdot 10^{32}\text{W}. \end{aligned} \tag{10.111}$$

The luminosity of the sun is

$$P_\odot = 3.85 \cdot 10^{26}\mathrm{W}.$$

The pulsar radiates about 10^6 times as much power as the sun.

5.3 The gravitational pressure is

$$\begin{aligned} P_\mathrm{g} &= \tfrac{F_\mathrm{g}}{4\pi R_0^2} = \tfrac{G\cdot M(R)}{4\pi R_0^2} \cdot \int_{R_0}^{\infty} 4\pi r^2 \varrho(r)\mathrm{d}r \\ &= G \cdot \tfrac{4}{3}\pi\varrho_0 R_0 \int_{R_0}^{\infty} r^2 \cdot \mathrm{e}^{-ar}\mathrm{d}r \\ &= \tfrac{4}{3}\pi\varrho_0 G R_0 \cdot \mathrm{e}^{-aR_0}\left[\tfrac{R_0^2}{a} + \tfrac{2R_0}{a^2} + \tfrac{2}{a^3}\right]. \end{aligned} \tag{10.112}$$

For the example of our sun the numerical values are

$$\begin{aligned} &\varrho_0 = 1.5 \cdot 10^3 \mathrm{kg/m^3}, R_0 = 7 \cdot 10^8 \mathrm{m}, a = 6 \cdot 10^{-8}\mathrm{m}^{-1} \\ &\Rightarrow P_\mathrm{g} = \tfrac{4}{3}\pi \cdot 1.5 \cdot 10^3 \cdot 6.67 \cdot 10^{-11} \cdot 7 \cdot 10^8 \cdot \mathrm{e}^{-42} \\ &\qquad \cdot\left[8.2 \cdot 10^{24} + 0.4 \cdot 10^{24} + 10^{22}\right]\mathrm{Pa} \\ &\qquad = 3 \cdot 10^2 \cdot \mathrm{e}^{-42} \cdot 8.6 \cdot 10^{24}\mathrm{Pa} \\ &\qquad \approx 1.5 \cdot 10^9 \mathrm{Pa} \approx 1.5 \cdot 10^4 \mathrm{atm}. \end{aligned} \tag{10.113}$$

5.4 The relative frequency-shift is, according to (5.104)

$$\frac{\Delta\nu}{\nu} = \frac{\mathcal{R}}{2R}$$

with the Schwarzschild radius

$$\begin{aligned} \mathcal{R} &= \tfrac{2GM}{c^2} \\ \Rightarrow \tfrac{\Delta\nu}{\nu} &= \tfrac{G\cdot M}{M\cdot c^2}. \end{aligned}$$

(a) For the earth- surface is $M \approx 6 \cdot 10^{24}\mathrm{kg}, R = 6.36 \cdot 10^6\mathrm{m}$

$$\Rightarrow \frac{\Delta\nu}{\nu} = \frac{6.67 \cdot 10^{-11} \cdot 6 \cdot 10^{24}}{6.36 \cdot 10^6 \cdot 9 \cdot 10^{16}} = 7 \cdot 10^{-10}. \tag{10.114}$$

For $\nu = 6 \cdot 10^{14}\mathrm{s}^{-1}$ (red light) is

$$\Delta\nu = 4 \cdot 10^5 \mathrm{s}^{-1} = 0.4\,\mathrm{MHz}.$$

(b) For the neutron star is

$$\begin{aligned} \tfrac{\Delta\nu}{\nu} &= \tfrac{6.67\cdot 10^{-11}\cdot 4\cdot 10^{30}}{10^4\cdot 9\cdot 10^{16}} = 3 \cdot 10^{-1} \\ \Rightarrow \Delta\nu &= 1.8 \cdot 10^{14}\mathrm{s}^{-1} \quad \text{für} \quad \nu = 6 \cdot 10^{14}\mathrm{s}^{-1}. \end{aligned} \tag{10.115}$$

5.5 For a density ϱ and a mean atom mass number A of atoms with the mass $m = A \cdot 1\,\mathrm{AMU}$ the mean distance d between two atoms is with

$$\begin{aligned} &\varrho = \frac{A \cdot 1\mathrm{AMU}}{d^3}, 1\mathrm{AMU} = 1.66 \cdot 10^{-27}\mathrm{kg} \\ &\Rightarrow d = \sqrt[3]{\frac{50 \cdot 1.66 \cdot 10^{-27}}{4 \cdot 10^3}}\mathrm{m} \\ &= 2.75 \cdot 10^{-10}\mathrm{m}. \end{aligned} \tag{10.116}$$

The total number of atoms with mass m in a sphere with radius R is

$$N = \frac{\frac{4}{3}\pi R^3}{d^3} = \frac{4}{3}\pi R^3 \cdot \frac{\varrho}{m}. \tag{10.117}$$

The total binding energy is then

$$E_\mathrm{B} = -\frac{1}{2}N \cdot \varepsilon = \frac{2}{3}\pi R^3 \varrho \cdot \varepsilon/m.$$

where ε is the binding energy between two atoms.

The gravitational energy is (see problem 4.13)

$$E_\mathrm{grav} = -\frac{3}{5}GM^2/R = -\frac{3}{5}GN^2m^2/R. \tag{10.118}$$

For $\left|E_\mathrm{grav}\right| > |E_\mathrm{B}|$ is

$$\frac{3}{5}GN^2m^2/R > \frac{2}{3}\pi R^3 \varrho\varepsilon/m. \tag{10.119}$$

Inserting $N = \frac{4}{3}\pi R^3\varrho/m$ gives

$$\begin{aligned} R &> \sqrt{\tfrac{10\varepsilon}{16\pi\cdot m\cdot G\cdot\varrho}} = 1.6 \cdot 10^6\mathrm{m} = 1600\,\mathrm{km} \\ \Rightarrow M &> \tfrac{4}{3}\pi\varrho R^3 = 6.86 \cdot 10^{22}\mathrm{kg}. \end{aligned} \tag{10.120}$$

This is one order of magnitude smaller than the mass of the earth ($M_\mathrm{E} = 6 \cdot 10^{23}\mathrm{kg}$).

5.6 If the neutron star with mass M and radius R rotates with the angular velocity ω, the centrifugal force on a mass m at the equator is

$\mathrm{F_c} = \mathrm{m} \cdot \omega^2 \cdot \mathrm{R}$

The gravitational force acting on m is

$\mathrm{F_G} = \mathrm{G} \cdot \mathrm{m} \cdot \mathrm{M/R^2}$

The body m will be ejected, if $\mathrm{F_c} > \mathrm{F_G}$.

$\rightarrow \omega > \sqrt{\left(\mathrm{G} \cdot \mathrm{M/R^3}\right)}$

Inserting the numerical values $\mathrm{G} = 6.67\cdot 10^{-11}$; $\mathrm{M} = 1.5\ \mathrm{M_\odot} = 3\cdot 10^{30}$ kg, $\mathrm{R} = 10$ km $= 10^4$ m

$\rightarrow \omega = 1.4 \cdot 10^4\mathrm{s}^{-1} \rightarrow \mathrm{T} = 2\pi/\omega = 4.5 \cdot 10^{-4}\mathrm{s} = 0.45$ ms.

5.7 The neutrinos are produced by the reaction

$\mathrm{e^-} + \mathrm{p^+} \rightarrow \mathrm{n} + \nu_\mathrm{e}$ and

$\mathrm{e^+} + \mathrm{n} \rightarrow \mathrm{p^+} + \overline{\nu}_\mathrm{e}$.

They are detected by the inverse reactions

$$\begin{aligned} \nu_\mathrm{e} + \mathrm{n} &\rightarrow \mathrm{p^+} + \mathrm{e^-} \\ Q &= (m_\mathrm{n} - m_\mathrm{H})c^2 = 0.782\,\mathrm{MeV} \end{aligned}$$

and

$$\begin{aligned} \overline{\nu}_\mathrm{e} + \mathrm{p^+} &\rightarrow \mathrm{n} + \mathrm{e^+} \\ Q &= \left(m_\mathrm{p} - m_\mathrm{n} - m_\mathrm{e}\right)c^2 = -1.80\,\mathrm{MeV} \end{aligned}$$

and

$$\nu_\mathrm{e} + \mathrm{e^-} \rightarrow \nu_\mathrm{e} + \mathrm{e^-}.$$

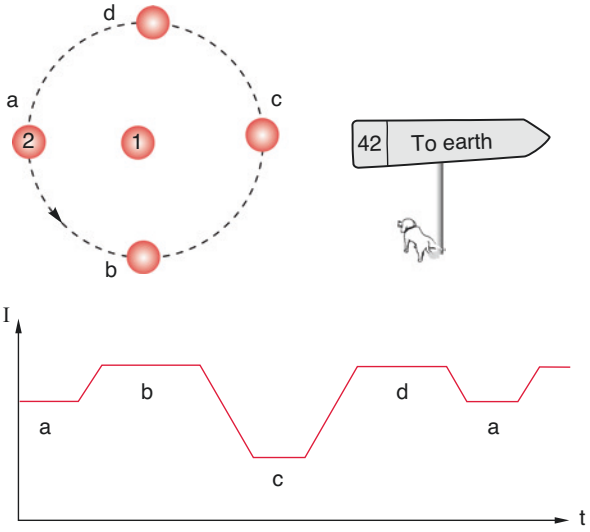


Fig. 10.6 To solution 5.1

Measured detection rate:
$\Phi\sigma N_p = \frac{dN}{dt} = 10/10\,\mathrm{s} = 1/\mathrm{s}$
The number of incident neutrinos can be obtained as follows:

$$\sigma \approx 4 \cdot 10^{-46}\mathrm{m}^2;$$

$$\begin{aligned}
N_p &= 2N_{H_2O} = 2 \cdot \frac{m_{H_2O}}{M_{molar}} \cdot N_A \\
&= 2 \cdot \frac{2.14 \cdot 10^6 \mathrm{kg}}{18\,\frac{\mathrm{g}}{\mathrm{mol}}} \cdot 6.022 \cdot 10^{23}\mathrm{mol}^{-1} \\
&= 1.432 \cdot 10^{32} \\
\Rightarrow \Phi &= \frac{\dot{N}}{\sigma \cdot N_p} = \frac{1\mathrm{s}^{-1}}{4 \cdot 10^{-46}\mathrm{m}^2 \cdot 1.432 \cdot 10^{32}} \\
&= 1.746 \cdot 10^{13}\mathrm{m}^{-2}\mathrm{s}^{-1} (\text{nur in } \Delta t = 10\mathrm{s}) \\
D &= 50 \cdot 10^3 \cdot 3.0856778 \cdot 10^{16}\mathrm{m}^2 \\
&= 1.54284 \cdot 10^{21}\mathrm{m}, \\
N_{\nu_e} &= 4\pi D^2 \cdot \Phi \cdot \Delta t \\
&= 4\pi \cdot 50^2 \cdot 10^6 \cdot 3.08567758^2 \cdot 10^{32}\mathrm{m}^2 \\
&\quad \cdot 1.746 \cdot 10^{13}\,\mathrm{m}^{-2}\mathrm{s}^{-1} \cdot 10\,\mathrm{s} \\
&= \pi \cdot 3.085678^2 \cdot 1.746 \cdot 10^{56} \\
&= 5.223 \cdot 10^{57}.
\end{aligned} \tag{10.121}$$

All three neutrino types (ν_e, ν_μ, ν_τ) are emitted in about equal rates, but only the type ν_e is measured. The total energy of the three neutrino- types is then

$$\begin{aligned}
E_{total} &= 3 \cdot N_{\nu_e} \cdot \overline{E} \\
&= 3 \cdot 5.223 \cdot 12 \cdot 10^{57}\mathrm{MeV} \\
&= 3.0126 \cdot 10^{46}\mathrm{J}.
\end{aligned}$$

$(1\,\mathrm{eV} = 1.6022 \cdot 10^{-19}\mathrm{As\,V} = 1.6022 \cdot 10^{-19}\mathrm{J}.)$

Chapter 6

6.1 The mean capture time of neutrons by protons is for a given capture cross section σ given by

$$\bar{t} = \frac{1}{n_p \cdot \sigma \cdot \bar{v} \cdot \sqrt{2}}, \tag{10.122}$$

where

$$\begin{aligned}
n_p &= \frac{\varrho_p}{m_p} \\
&= \frac{2 \cdot 10^{-3}\mathrm{kg/m}^3}{1.67 \cdot 10^{-27}\mathrm{kg}} = 1.2 \cdot 10^{24}\mathrm{m}^{-3}
\end{aligned} \tag{10.123}$$

is the proton density and $\sigma = 10^{-32}\mathrm{m}^2$.
At the temperature T is the mean relative velocity

$$\bar{v}_r = \bar{v} \cdot \sqrt{2} \quad \text{with} \quad \bar{v} = \left(\frac{8kT}{\pi \cdot m}\right)^{1/2}.$$

Inserting the numerical values gives

$$\begin{aligned}
\bar{v}_r &= \left(\tfrac{16kT}{\pi \cdot m}\right)^{1/2} \\
&= \left(\frac{16 \cdot 1.38 \cdot 10^{-23} \cdot 10^9}{\pi \cdot 1.67 \cdot 10^{-27}}\right)^{1/2}\mathrm{m/s} \\
&= 6.5 \cdot 10^6\mathrm{m/s} \\
\Rightarrow \bar{t} &= \frac{1}{1.2 \cdot 10^{24} \cdot 10^{-32} \cdot 6.5 \cdot 10^6}\mathrm{s} \\
&= 1.3 \cdot 10^1\,\mathrm{s} = 13\,\mathrm{s}.
\end{aligned} \tag{10.124}$$

Compared with the decay of free neutrons ($\tau = 800\,\mathrm{s}$) the capture is more probable by the factor 61.

6.2 At the expansion the number of photons remains constant. In the wavelength interval from λ to $\lambda + \mathrm{d}\lambda$ there are

$$\mathrm{d}N = \frac{16\pi^3 R^3}{h \cdot c/\lambda} \cdot \frac{hc^2}{\lambda^5} \cdot \frac{1}{\mathrm{e}^{hc/\lambda kT} - 1}\mathrm{d}\lambda \tag{10.125}$$

photons in the volume $V = 2\pi^2 R^3$ (see Eq. 6.14)

$$\Rightarrow N = \int_0^\infty \frac{16\pi^3 R^3 c}{\lambda^4} \frac{1}{\mathrm{e}^{hc/\lambda kT} - 1}\mathrm{d}\lambda. \tag{10.126}$$

Since $T = T_0 \cdot R_0/R$ and $\lambda = \lambda_0 \cdot R/R_0 \Rightarrow \mathrm{d}\lambda = (R/R_0)\mathrm{d}\lambda_0$

$$\begin{aligned}
&\Rightarrow \int_0^\infty \tfrac{16\pi^3 R^3 \cdot c}{\lambda^4} \tfrac{1}{\mathrm{e}^{hc/\lambda kT} - 1}\mathrm{d}\lambda \\
&= \int \tfrac{16\pi^3 R^3 c \cdot R_0^4}{\lambda_0^4 R^4} \cdot \tfrac{(R/R_0)\mathrm{d}\lambda_0}{\exp\left[\frac{hc \cdot R_0 \cdot R}{\lambda_0 \cdot R \cdot k \cdot T_0 \cdot R_0}\right] - 1} \\
&= \int \tfrac{16\pi^3 R_0 c}{\lambda_0^4} \tfrac{1}{\mathrm{e}^{hc/\lambda_0 kT_0} - 1}\mathrm{d}\lambda_0.
\end{aligned} \tag{10.127}$$

Since this is valid for arbitrary values of T_0 the integrands on both sides of the equation must be equal, i.e. the distribution in the expanded universe is again a Planck- distribution but at a lower temperature.

6.3 We compose the energy density w as the sum

$$W = \varrho_{\mathrm{m}} c^2 + w_{\mathrm{rel}}$$

of mass-energy density of the nonrelativistic matter (baryons and electrons) and the energy density of the relativistic particles (Neutrinos and photons).

The pressure of the nonrelativistic particles

$$p_{\mathrm{m}} = \frac{1}{3}\varrho_{\mathrm{m}}\overline{v^2} = \frac{1}{3}w_{\mathrm{m}}\overline{v^2}/c^2$$

is with $v \ll c$ negligible against the pressure

$$p_{\mathrm{rel}} = \frac{1}{3}\varrho_{\mathrm{rel}} \cdot c^2 = \frac{1}{3}w_{\mathrm{rel}}$$

of the relativistic particles

For the non- relativistic particles we obtain with $p_{\mathrm{m}} \approx 0$

$$\begin{gathered} \tfrac{\mathrm{d}}{\mathrm{d}t}\left(\varrho_{\mathrm{m}} c^2 R^3\right) = 0 \\ \Rightarrow \varrho_{\mathrm{m}} c^2 R^3 = \text{const.} \\ \Rightarrow \varrho_{\mathrm{m}} \propto R^{-3}, \end{gathered}$$

While for the energy density of the relativistic particles

$$\begin{gathered} \tfrac{\mathrm{d}}{\mathrm{d}t}\left(w_{\mathrm{rel}} \cdot R^3\right) = -\tfrac{1}{3} w_{\mathrm{rel}} \cdot \tfrac{\mathrm{d}}{\mathrm{d}t}\left(R^3\right) \\ \Rightarrow R^3 \cdot \tfrac{\mathrm{d}w_{\mathrm{rel}}}{\mathrm{d}t} + w_{\mathrm{rel}} \cdot 3R^2 \cdot \tfrac{\mathrm{d}R}{\mathrm{d}t} = -w_{\mathrm{rel}} R^2 \tfrac{\mathrm{d}R}{\mathrm{d}t} \\ \Rightarrow R^3 \tfrac{\mathrm{d}w_{\mathrm{rel}}}{\mathrm{d}t} = -4 w_{\mathrm{rel}} R^2 \tfrac{\mathrm{d}R}{\mathrm{d}t}. \end{gathered} \tag{10.128}$$

Integration gives

$$\begin{gathered} 4\ln R = -\ln w_{\mathrm{rel}} + C \\ \Rightarrow \ln\left(R^4/w_{\mathrm{rel}}\right) = C \\ \Rightarrow R^4/w_{\mathrm{rel}} = \mathrm{e}^C = C^* = \text{const.} \\ \Rightarrow w_{\mathrm{rel}} \propto 1/R^4. \end{gathered} \tag{10.129}$$

6.4 In case of a uniform distribution of galaxies with the absolute magnitude M the number of observed galaxies with apparent magnitude $\geq$ m is

$$M_m = \frac{4}{3}\varrho\pi r_m^3$$

where ϱ is the density of galaxies which is assumed to be constant.

The received energy flux density ϕ of a galaxy with absolute magnitude M is

$$\phi \propto 1/r^2$$

The energy flux density of a galaxy at the distance r_m is therefore larger by the factor $(r_{m+1}/r_m)^2$ than that of a galaxy with the same absolute magnitude at the distance r_{m+1}.

The ratio of the energy flux densities ϕ of two galaxies with apparent magnitudes m resp. m + 1 are equal to the ratio

$$\phi_m/\phi_{m+1} = 10^{0.4}$$

Therefore, we get with $(r_{m+1}/r_m)^2 = 10^{0.4}(r_{m+1}/r_m)^2 = 10^{0.4}$

$$\frac{N_{m+1}}{N_m} = \left(\frac{r_{m+1}}{r_m}\right)^3 = \left(10^{0.4}\right)^{3/2} = 10^{0.6} \approx 4.$$

6.5 The space-time-metric is determined by

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}l^2$$

where the line element dl is given by Eq. (6.18). The world line of a photon travelling with the velocity of light c is determined by $\mathrm{d}s = 0$

$$\Rightarrow (\mathrm{d}l/\mathrm{d}t)^2 = c^2 \Rightarrow \mathrm{d}l/\mathrm{d}t = c.$$

The distance of a galaxy is proportional to the scale parameter R(t)

$$R(t) = (R(t)/R_0) \cdot r_0.$$

The line element dl is for a flat universe

$$\begin{gathered} \mathrm{d}l = R \cdot \mathrm{d}r \\ \Rightarrow \frac{\mathrm{d}r}{\mathrm{d}t} = \frac{1}{R}\frac{\mathrm{d}l}{\mathrm{d}t} = \frac{1}{R} \cdot c = \frac{c}{R_0}\left(\frac{t_0}{t}\right)^{2/3} \end{gathered}$$

because for a flat universe is (see Eq. 6.71)

$$R/R_0 = (t/t_0)^{2/3}.$$

Integration from von $r_{\mathrm{c}} = r(t_{\mathrm{c}})$ (time of the emission of the photon) until $r(t_0)$ (the present distance) gives

$$\begin{aligned} \int_{t_{\mathrm{e}}}^{t_0} \tfrac{\mathrm{d}r}{\mathrm{d}t}\mathrm{d}t &= \int_{r_{\mathrm{e}}}^{r_0} \mathrm{d}r = \tfrac{c}{R_0}\int_{t_{\mathrm{e}}}^{t_0}\left(\tfrac{t_0}{t}\right)^{2/3}\mathrm{d}t \\ &= \tfrac{3c}{R_0} t_0^{2/3} \cdot t^{1/3}\Big|_{t_{\mathrm{e}}}^{t_0} \\ r_0 - r_{\mathrm{e}} &= \tfrac{3c}{R_0}\left[t_0 - t_0^{2/3} t_{\mathrm{e}}^{1/3}\right] \\ &= \tfrac{3ct_0}{R_0}\left[1 - \left(\tfrac{t_{\mathrm{e}}}{t_0}\right)^{1/3}\right]. \end{aligned} \tag{10.129}$$

According to Eq. (6.82) is

$$\begin{gathered} (t_0/t_{\mathrm{e}})^{2/3} = 1 + z \\ \Rightarrow \left(\tfrac{t_{\mathrm{e}}}{t_0}\right)^{1/3} = \tfrac{1}{\sqrt{1+z}} \\ t_0 = \tfrac{2}{3}/H_0 \\ \Rightarrow r_0 - r_{\mathrm{e}} = \tfrac{2c}{H_0}\left[1 - \tfrac{1}{\sqrt{1+z}}\right]. \end{gathered} \tag{10.130}$$

The present distance r(z) of a galaxy which emits light of the wavelength λ, which appears at the detector redshifted by $\Delta\lambda$ is with $z = \Delta\lambda/\lambda$

$$r_0(z) = r_e + \frac{2c}{H_0}\left[1 - \frac{1}{\sqrt{1+z}}\right]. \qquad (10.131)$$

For large values of z is $r_e \ll r_0$ and we obtain the Eq. (6.83). With $t_0 = t_e \cdot (1+z)^{3/2}$ it follows for the distance the light has travelled between emission time and detection time

$$\begin{aligned} L(z) &= c \cdot (t_0 - t_e) = c \cdot t_0\left(1 - \tfrac{1}{(1+z)^{3/2}}\right) \\ &= \tfrac{2C}{3H_0}\left(1 - \tfrac{1}{(1+z)^{3/2}}\right), \end{aligned} \qquad (10.132)$$

because $t_0 = 2/(3H_0)$.

6.6 The absolute magnitude M of stars is related to their apparent magnitude m by

$$M = m - 5 \cdot \log\left(\frac{r}{10\,\mathrm{pc}}\right) - A(r)$$

where A(r) is the absorption of the star radiation on its way through the interstellar matter to the earth.

If $\phi(M)$ is the luminosity of stars with the absolute magnitude M and D is the density of stars in the distance interval from r to $r + \mathrm{d}r$ and in the solid angle interval $\mathrm{d}\Omega$ in the direction of the galactic length l and the galactic latitude b, we get the number of stars with the apparent magnitude m in the volume element $\mathrm{d}V = r^2 \cdot \mathrm{d}\Omega \cdot \mathrm{d}r$

$$\mathrm{d}N(m) \cdot \mathrm{d}\Omega = D(r, l, b, \phi(M))\mathrm{d}V \qquad (10.133)$$

with

$$\phi(M) = \Phi\left(m - 5 \cdot \log\frac{r}{10\,\mathrm{pc}} - A(r)\right).$$

Integration over all distances r gives the number

$$\begin{aligned} N(m)\mathrm{d}\Omega &= \int_0^\infty D(r, l, b, \phi(M))r^2\mathrm{d}r\mathrm{d}\Omega \\ &= \int_0^\infty D\big[r, l, b, \Phi(m - 5 \cdot \log(r/10\,\mathrm{pc}) \\ &\qquad -A(r)))]r^2\mathrm{d}r \cdot \mathrm{d}\Omega. \end{aligned} \qquad (10.134)$$

The numerical solution of this expression gives the density $D(r, l, b)$ from the measured number N(m) for different apparent magnitudes m in the different directions (l, b) of the sky. The total density of stars is then obtained by integration

$$D(r, l, b) = \int_0^\infty D(r, l, b, \phi(M))\mathrm{d}\phi.$$

Chapter 7

7.1 For a uniform distribution of galaxies with the absolute magnitude M the number of galaxies in the radial interval dr at the distance r_m is

$$N_M = 4\pi\rho r^2 \mathrm{d}r$$

The intensity of the radiation from the galaxy with magnitude M decreases with $1/r^2$.

The observed intensities of galaxies at the distances r_m resp. r_{m+1} is proportional to $(r_{m+1}/r_m)^2$

According to the relation between intensity and magnitude

$$m = -2.5 \cdot \log_{10}(\Phi/\Phi_0)$$

we obtain

$$\Phi_m/\Phi_{m+1} = 10^{0.4}/1 \rightarrow (r_{m+1}/r_m)^2 = 10^{0.4}$$

We therefore obtain

$$N_{m+1}/N_m = (r_{m+1}/r_m)^3 = \left(10^{0.4}\right)^{3/2} = 10^{0.6} \approx 4.$$

7.2 From Fig. 10.7 we get the following relations:

$$\cos\alpha = (R_0/R) \cdot \sin l$$

$$R/R_0 = \sin l/\sin(\pi/2 + \alpha)$$

We then obtain for the velocit components of a star the conditions

$$\begin{aligned} v_r &= v \cdot \cos\alpha - v_0 \cdot \sin(l) \\ &= \omega \cdot R \cdot (R_0/R)\sin(l) - \omega_0 \cdot R_0 \cdot \sin(l) \\ &= (\omega - \omega_0)R\cos\alpha \\ V_t &= R_0(\omega - \omega_0) \cdot \cos(l) - \omega \cdot r \end{aligned}$$

Since the radial change of $\omega(r)$ is small, we can expand ω into a Taylor series around ω_0 which gives.

$\omega = \omega_0 + \Delta R \cdot \mathrm{d}/\mathrm{d}R|_R = R_0 + \ldots.$ Inserting into the equation for v_t gives

$$V_t = r_0 \Delta R \cdot \mathrm{d}v/\mathrm{d}R \cdot \cos(l) - \omega \cdot r$$

For stars close to the sun is

$$\Delta R = -r \cdot \cos(l)$$

We then obtain

$$\omega \approx \omega_0 - r \cdot \mathrm{d}\omega/\mathrm{d}R = R_0 \cdot \cos(l)$$

This gives finally

$$V_t \approx A \cdot r \cdot \cos(2l) + B$$

where the Oort's constants are

$A = 1/2(v_0/R_0 - dv/dR)_{R0}$

$B = -1/2(v_0/R_0 + dv/dR)_{R0}$

7.3 If we assume that all stars move on circles around the galactic center the Doppler broadening is zero. The broadening mechanism is therefore only pressure broadening which increases with increasing density. The line profile is then a Lorentzian profile

$$\alpha(\omega) = \alpha 0/(\alpha - \alpha 0)^2 + \gamma)$$

7.4 We start with the relation

$$m - M = 5 \cdot \log_{10}(r/10pc)$$

$$\rightarrow r = 10pc \cdot 10^{(m-M)/5}$$

With m = 14.16 and M = 0

$$\rightarrow r = 10pc \cdot 10^{2.83}$$

$$= 10pc \cdot 680 = 6800pc = 22.200ly$$

The distance of M13 is r = 22.200 ly.
Its diameter is D = r · α with α = 25 ' = 25/3600 rad =

$$\rightarrow D = 22.200 \cdot 25/3600 = 154\,ly = 48\,pc$$

The absolute magnitude of M13 is obtained from

$$m - M = 5 \cdot \log_{10}(r/10pc)$$

$$\rightarrow M = m - 5 \cdot \log 680 = 5.8 - 5 \cdot 2.8 = -8.2$$

With M_1(single star) = 0 and M_2(M13) = -8.2 we get the number N of stars in M 13 from the ratio of the luminosities L for the whole cluster and L_2 for a single star.

$$M_1 - M_2 = -2.5 \cdot \log_{10}(L/N \cdot L) = 2.5 \cdot \log N$$

$$\rightarrow N = 10_{(M1-M2)/2.5} = 10^{8\cdot 2/2.5} = 10^{3\cdot 28} = 1900.$$

In fact the cluster contains about 500.000 stars.

The variable star used here as example, has a much larger luminosity than the average star. Therefore, the number N comes out as too low.

Chapter 8

8.1 The gravitational force acting onto a mass m at the distance r from the center is determined by the mass M in a sphere with radius r. The mass in the outer volume with radius >r acts with the total force F = 0 onto m.

Point-like mass M at r =0

Gravitational force = centripetal force:

$$G \cdot mM/r^2 = m \cdot \omega^2 \cdot r$$

$$\rightarrow \omega^2(r) = G \cdot M/r^3 \rightarrow v(r) = \omega \cdot r = (G \cdot M/r)^{1/2}.$$

The angular velocity decreases with increasing radius r as $1/r^{3/2}$, the velocity as $1/r^{1/2}$

Uniform mass distribution with mass density ρ

$$M = (4/3) \cdot \pi r^3 \cdot \rho$$

$$G \cdot (4/3)\pi \cdot r^3 \cdot \rho/r^2 = \omega^2 \cdot r$$

$$\rightarrow \omega^2 = (4/3)\pi \cdot \rho$$

The angular velocity is independent of r inside the uniform mass distribution!

$$\boldsymbol{\rho(r) = \rho_0 \cdot (1 - r/R)}$$

The mass in a sphere from r = = until r > 0 is now

$$M(r) = (4/3)\pi r^3 \cdot \rho_0(1 - r/R)dr$$

The angular velocity ω(r) can be obtained as before from

$$G \cdot M(r)/r^2 = \omega^2(r) \cdot r$$

$$\rightarrow \omega^2 = (4/3)G \cdot \pi\rho_0(1 - r/R).$$

The angular velocity ω decreases from

$$\omega(r = 0) = \left((4/3)G\pi\rho_0\right)^{1/2}$$

to

$$\omega(r = R) = 0.$$

The velocity

$$\rightarrow v^2 = (4/3)G \cdot \pi\rho_0\left(r^2 - r^3/R\right)$$

The velocity is v(r = 0) = 0 and v(r = R) = 0.
The maximum velocity is obtained for

$$dv^2/dr = 0 \rightarrow 2r_m - 3r_m^2/R = 0$$

$$\rightarrow r_m = (2/3) \cdot R.$$

8.2 The expansion velocity of a galaxy is retarded by the baryonic matter and accelerated by the dark energy. For a homogeneous mass distribution with density ρ_0 the equation of motions is

$$M \cdot d^2R/dt^2 = 4/3)\pi \cdot R^3 \cdot \left(-\rho_0/R^2 + a_{DE}\right).$$

For $\rho_0/R^2 > a_{DE}$ the expansion is decelerated, for $\rho_0/R^2 < a_{DE}$ it is accelerated.

Chapter 9

9.1 The effective potential energy of a rotating mass is

$$E_{eff} = E_{pot} + E_{rot}.$$

Here the potential gravitation energy of the gaseous (spherical) mass M with radius r is

$$E_{pot} = -3G \cdot M^2/5r$$

while the rotational energy is

$$E_{rot} = 1/2 \cdot I \cdot \omega^2 = 1/2 \cdot L^2/I,$$

where I is the moment of inertia and $L = I \cdot \omega$ the angular momentum.

The moment of inertia of a spherical mass with radius r is

$$I = 2/5 \cdot M \cdot r^2$$

The rotational energy is then

$$E_{rot} = +(5/4)L^2/\left(Mr^2\right).$$

For a given angular momentum L $E_{ro\ t}$ depends only on r. Therefore, E_{rot} can be regarded as part of the effective potential energy, which is then written as

$$\begin{aligned} E^{eff} &= -3G \cdot M^2/5r + (5/4)L^2/\left(Mr^2\right) \\ &= -3G \cdot M^2/5r + (5/4)I^2\omega^2/M \cdot r^2). \end{aligned}$$

Inserting $M = (4/3)\pi\rho r^3$ and $I = (2/5)Mr^2 =$ we obtain

$$\begin{aligned} E^{eff} &= -4/5)\pi G \cdot \rho \cdot r^2 + (8/15)\pi \cdot \rho \cdot r^5 \cdot \omega^2 \\ &= (4/5)\pi \cdot \rho\left(-G \cdot r^2 + (1/3)r^5\omega^2\right). \end{aligned}$$

A minimum occurs for $dE/dr = 0$. This gives the relation

$$\begin{aligned} &(8/5) \cdot \pi \cdot G \cdot \rho \cdot r = (8/3)\pi \cdot \rho \cdot r^4\omega^2 \\ &\rightarrow r_{min} = \left((3/5) \cdot \left(G/\omega^2\right)^{1/3}\right. . \end{aligned}$$

The effective potential energy at r_{min} is then

$$\begin{aligned} E^{eff}(r_{min}) &= \left(G^{5/3}/\omega^{4/3}\right) \cdot \left(-36 \cdot \pi \cdot \rho + (9/5)^2/\omega^2\right))/125 \\ &= \text{Potential depth.} \end{aligned}$$

For $r \rightarrow \infty$ the effective potential approaches $E^{eff} \rightarrow 0$.

9.2 Calculate the increase Δh of the ocean height h per degree temperature increase for an average ocean depth of 3800 m, when the volume expansion is $\Delta V = \Delta h \square \Delta A$.

Volume expansion coefficient of water at $T = 15^0$ C is

$$\begin{aligned} &\alpha = \Delta V/V = 1.5 \cdot 10^{-4}/K; \rightarrow \Delta h/h \\ &= (1/3) \cdot \Delta V/V = 5 \cdot 10^{-5}/K. \end{aligned}$$

If we assume, that the whole water down to the ground participates equally in the temperature increase, we get for the ocean depth $h = 3800$ m an increase of the water level by

$$\begin{aligned} \Delta h &= 3800 \cdot 5 \cdot 10^{-5} m/K \\ &= 1.9 \cdot 10^{-1} m/k. \end{aligned}$$

A temperature increase of 2 K would increase the ocean level by about 38 cm.

9.3 Calculate the change of the potential energy, if 10^9kg of material with drop in the early earth ($M = 5 \square 10^{24}$ kg) from r = 6000 km to r = 1000 km, if the density $\rho(r)$ of the earth follows the function $\rho(r) = \rho_0(1\text{-}r/R)$ with $\rho_0 = 11$ kg/dm^3.

The gravitational potential energy of a mass m at the distance r from the center is

$$E_{pot} = -G \cdot M(r) \cdot m/r.$$

The difference $\Delta E_{pot} = E_{pot}(m, r_1) - E_{pot}(m, r_2)$ is

$$G \cdot m[M(r_1)/r_1 - M(r_2)/r)].$$

With $M(r) = (4/3)\pi\rho r^3 = (4/3)\pi \cdot \rho_0(1 - r/R) \cdot r^3$ we get

$$\Delta E_{pot} = G \cdot (4/3)\pi\rho_0 \cdot m \cdot \left[(r_2/R) \cdot r_2^2 - r_2^2 - (r_1/R)r_1^2 + r_1^2\right].$$

Inserting the numerical values

$$\begin{aligned} &G = 6.67 \cdot 10^{-11} m^3 kg^{-1} s^{-2}, m = 10^9 kg, \\ &M(r) = \left(r^3/R^3\right) \cdot M_0 \text{with } M_0 = 5 \cdot 10^{24} kg, \\ &\rho_0 = 11 \cdot 10^3 kg/m^3, r_1 = 1 \cdot 10^6 m, r_2 = 6 \cdot 10^6 m \end{aligned}$$

we get,

$$\begin{aligned} &\Delta E_{pot} = 6.67 \cdot 10^{-11} \cdot (4/3)\pi \cdot 11 \cdot 10^3 \cdot 10^9 \\ &\left[6^3 \cdot 10^{18}/6.38 \cdot 10^6 - 10^{18}/6.38 \cdot 10^6 - 35 \cdot 10^{12}\right] N \cdot m \\ &= 307 \cdot 10 \cdot 1.35 \cdot 10^{12} N \cdot m = 4.15 \cdot 10^{15} N \cdot m. \end{aligned}$$

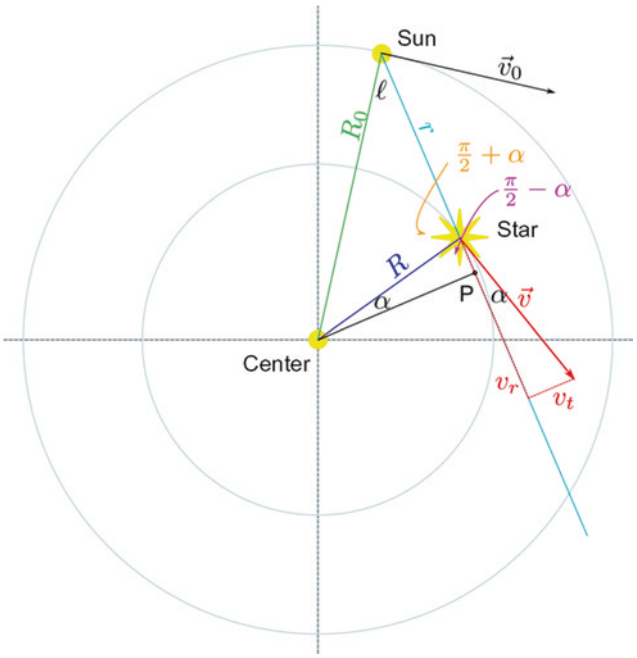


Fig. 10.7 To solution 7.2 (Ulm University, Othmar Marti) http://wwwex.physik.uni-ulm.de/lehre/kta-2018-2019/KTA-2018-2019se58.html

Time Table of Astronomy

35.000 BEC During the early Palelythic period people followed the moon phases by scratching lines in animal bones.

4.900 BEC Gosec Circle is probably the worldwide oldest sun observatory. It consists of a central ditch and 2 palisade rings with entrances in the direction of sun rise and sun set.

4th Millenium BEC. Egyptian calendar using a solar year.

with 365 days.

3rd Millenium BEC First systematic sky observations in.

Babylon, Egypt and Middle America.

with the invention of star constellations.

2500 BEC Stonhenge as sun observatory and place for celestial.

Observations.

2137 First Chinese report of a solar eclipse.

2000 BEC First quantitative measurements of planet motions in Babylon.

in Babylon.

1600 BEC Sky-disc of Nebra, Germany shows for the first time the sky.

time the sky with sun and 32 golden stars including the 7.

stars of the Plejades.

9th Century BEC. Description of star constellations in Chinese publication.

Publications.

from 700 BEC First systematic documentation of.

measurements, comparison of lunar.and solar calendar.

500–400 BEC culmination of Babylonian astronomy.

450 BEC The Greek philosopher Anaxagoras suggests that the stars are suns like our own sun.

the stars are suns like our own sun.

350 BEC. *Eudoxus*, Greek mathematician and astronomer introduces the geocentric model where the celestial bodies move on homocentric orbits around the earth.

300 BEC *Aristotl*e (384–322 b. Chr.) supports the geocentric philosophy and estimates the distances of sun and moon.

260 BEC. *Aristarch* from Samos, a Greek Philosopher supported as the first a helio-centric system.

240 BEC *Erathostenes* greek scientist from Kyrene postulates the spherical form of the earth. He measures the earth circumference as 46.660 km, which is no too far away from the present value of 40.000 km.

150 BEC *Hipparcos* (190–125 BEC) was the most famous and important astronomer in ancient history. He was the founder of the scientifiv astronomy, based on observations. He supported the geocentric theory and introduced the epicycles in order to explain the observed orbits of the planets. He measured the length of the solar year with an accuracy of 6 min.

100-170 AD *Ptoleme,* Greek mathematician and astronomer gives in his bool Almagest "a summary of the astronomical knowledge of his time". He supports the geocentric theory and the epicycles of Hypparcos whom he admires.

During the next 600 years not much progress in Astronomy has been achieved.

829 a.Chr. Foundation of the astronomical observatory in Bagdad.

1543 Publication of "*De Revolutionibus Orbium Coelestium*" by *Nicolaus Copernicus* where the heliocentric system was proposed based on stronger evidence than had been available to Aristrach.

1546–1601 *Tycho deBrahe*. Dansc astronomer, who observed a supernova 1572 in the milky way. He built several observatories with good observation tools. He measured the orbits of planets with high accuracy, but he rejected the system of Copernicus and developed a modified Ptoleme-system.

1571-1630 *Johannes Kepler* German astronomer and mathematician. He used the precise measurements of Tyche de Brahe and recognized that these data were not consistent with the geocentric epicyclic system. He published his first two laws in the book "Epitome astronomiae Copernicae" where he strongly defended the heliocentric system of Copernicus. In his following book Harmonices Mundi "he substantited the 3rd Kepler law". He observed 1604 a supernova explosion in the star constellation *Serpentarius.*

1582 Calendar reform by *Pope Gegor.*

1608 *Hans Lipershey*, a dutch spectacle maker develops the first lens telescope.

1610 *Galileo Galilei* (1564–1642) published his first astronomical observations with an improved Lippershey telescope. He discovered that the milky way consists of numerous individual stars, that there are mountains and

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics,* Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4

valleys on the moon and that Jupiter is orbitized by at least 4 moons (the Galileian moons), which falsified the theory that the planets were fixed on crystal spheres around the earth.

1616 The Catholic Church prohibits all books which support the heliocentric system.

1650 Discovery of the first binary (Mizar A and B) in the constellation Ursa Major by *Riccioli.*

1672 *Giovanni Cassini* determines the distances of the planets from the sun.

1675 Measurements of the speed of light by *Ole Römer* (1644–1710) based on the delayed occultation time of a Galileian Moon.

1687 *Isaac Newton* (1643–1727), English Astronomer, Mathematician and Physicist, published his gravitational law and the fundamental laws of the motion of bodies in the book "*Philosophiae Naturalis Prinicipia Mathematica*"

H*e* proposed a miror telescope instead of the lens telescope (Newton telescope).

1705 *Edmond Halley* (1656–1742) English astronomer, published his work on the orbits of comets and delivers for some comets precise orbital data.

1718 Halley discovers the proper motion of the stars Sirius, Aldebaran and Arkturus which he took as examples of the proper motion of all stars.

1719 *Laplace* proposes the concept of black holes.

1774 First determination of the earth mass by *Nevyl Makelyne* and independently by *Henry Cavendish* using a different method.

1781 *Friederich Wilhem Herschel* (1738–1822) discovers the planet Uranus.

1784 *Herschel* counts the number of stars and its dependence on the observation direction, He proposes the structure of our Milky Way.

Charles Messier publishes his catalogue of star clusters and nebulas.

1801 Discovery of the first planetoid Ceres by *G.Piazzi.*

1814 *Joseph von Fraunhofer* (1787–1826) builds a telescope with spectrometer and discovers dark lines in the solar spectrum (Fraunhofer lines).

1840 *Friedsrich Wilhelm Bessel*, Astronomer in Königsberg Germany, measures with high precision the parallaxes of stars and deduces from the temporal variation of the parallaxes of some stars that they must have a non-visible companion.

1846 *Johann Gottfried Galle* (1812–1910) discovers the planet Neptun based on the theoretical determination of its position by the French mathematician *U. Leverrier* calculated from perturbations of the orbit of Uranus.

1859 Foundation of the Spectral Analysis by Gustav Robert Kirchhoff (1824–1887) and *Robert Wilhelm Bunsen* (1811–1899).

1872 The American astronomer *Henry Draper* takes the first spectrum of a star.

1886 First application of spectral analysis to star spectra. Measurements of Doppler shifts of spectral lines by *W. Huggins* (1824–1910), English astronomer.

1901 *H.Draper* publishes an extensive catalogue of stars (Henry Draper Catalogue).

1910 *Williamina Fleming* publishes the first observation of a white dwarf.

1912 Discovery of the period- luminosity relation of variable Cepheide stars. Determination of larger distances. Proof that galaxies are systems of many stars outside our own galaxy.

1913 Hertzsprung Russel Diagramm by *Ejnar Hertzsprung* (1873–1967) *and Henry Norris Russel* (1877–1957).

1917 First light for the 2.5 m mirror telescope on Mount Wilson, at that time the largest telescope worldwide.

1924 With this telescope *Edwin Hubble* (1889–1953) could first resolve individual stars in neighbouring galaxies.

1925 Classification of galaxy types by E. Hubble.

1929 Measurements of the escape velocity of far distant galaxies by the red shift of spectral lines.

1930 Discovery of the radiation of the milky way with wavelengths in the meter range and foundation of radio-astronomy by *Karl Jansky* (1905–1950) and *G. Reber* (1911–2002).

1931 Calculation of the Chandrasekhar mass limit of white dwarfs by the Indian astronomer Subrahmanjan Chandrasekhar (1910–1995).

1938 Proposal of nuclear fusion processes in the central part of stars as energy source of stars by *Hans A. Bethe* (1906–2005) and *Carl Friederich von Weizsäcker* (1912–2007).

1939 Theory of black holes (*Oppenheimer et. al*).

1942 Verification of extragalactic radio sources.

1948 First light for the worldwide largest 5.1 m mirror telescope on the Mount Palomar in California.

1949 Investigation of the spiral structure of our galaxy by radio astronomy on the 21 cm line.

1950 First observation of molecular lines in the interstellar medium. Detection of many different molecules.

1957 The Soviet Union launched the first unmanned satellite Sputnik orbiting around the earth.

1958 Foundation of the NASA.

1963 Discovery of quasars by *Alan Rex Sandage* (1926–2010).

Marteen Schmidt measures the spectra of quasars.

1964 Discovery of the cosmic microwave radiation by *Arno Allen Penzias* and *Robert Woodrow Wilson.*

1967 Begin of the long baseline radio interferometry.

First discovery of a pulsar by *Jocelyn Bell* and *Antony.*

Hewish

1968 Proposal that pulsars are rotating neutron stars by *Thomas Gold.*

1969 First NASA mission that brought men (Neil Armstrong and Buzz Aldrin) to the moon.

1972 Initial operation of the worldwide largest mobile radio telescope with a 100 m parabolic mirror in Effelsberg Germany.

1976 The space probe Mariner 9 orbits around Mars.

1977 The spacecraft Voyager 1 was launched to explore the outer solar system.

1979 Alan Guth develops his model of the inflationary universe.

1982 Discovery of the gravitational lens effect for radiation from quasars passing nearby a galaxy.

1983 Nobelprize for S. Chandrasekhar for his star models which could explain white dwarfs.

The first infrared astronomical satellite IRAS is launched, which scanned about 98% of the sky before its supply of liquid helium was exhausted.

1986 The spaceprobe Giotto passes closely the comet Haley and sends pictures of the comet nucleus.

1987 observation and precise measurement of the radiation and the neutrinos of the supernova 1987aI in the Magellanian Cloud.

1989 Start of the astrometric satellite HIPPARCOS.

1990 Hubble Space telescope, discovery of the first extrasolar planet.

1992 The cosmic background explarer COBE measures the spatial distribution of the cosmic background radiation.

Introduction of adaptive optics for large earth bound.

telescopes.

1993 Nobelprize for Russel Hulse and Joseph H. Taylor for their discovery of a pulsar in a binary system.

1995 Start of the solar observatory satellite SOHO.

1997 NASA space probe pathfinder lands on the mars surface.

1998 Precise measurements of far distant supernova Ia as.

distance candles for the calibration of far distances.

First evidence of neutrino oscillations by the detector.

super-Kamiokande in Japan, solving the solar neutrino.

problem.

2002 Experimental confirmation of a black hole in the center of our galaxy with the ESO VLT on the Paranal in Chile., using infrared measurements of the position and velocities of stars close to the black hole by Genzel and Coworkers.

Combination of the two large 10 m Keck telescopes to form a large star interferometer.

Nobelprize to Raimond Davies and Masatashi Koshiba for the detection of solar neutrinos.

2003 NASA launched the satellite Spitzer orbiting around the sun following the earth orbit. It measures the composition of gas and dust in the interstellar space.

2004 First operation of the 600 m gravitational detector GEO 6oo close to Hannover in Germany.

2005 First interferometric combination of 4 large telescopes of the European spouthern observatory on the Paranal, allowing the angular resolution which corresponds to that of a 200 m telescope.

2006 Nobelprize to John C. Mather and George F. Smoot for their detailled measurements of the anisotropy of the cosmic background radiation.

Pluto looses its status as planet and is downgraded to a dwarf planet.

Upto 2006 more than 130 Exoplanets have been discovered.

2007 Verification that the star 555 Cancri in the constellation Cancer is orbited by more than 5 planets.

2011 Nobelprize for Saul Perlmutter, Brian Schmidt, Adam Ries for the discovery of the accelerated expansion of the universe.

2013 The Hubble Space Telescope detected the disintegration of a comet by sunlight.

2014 First detection of gravitational waves produced by the fusion of a neutron star with a black hole, observed by the LIGO cooperation.

2016 An exoplanet is discovered orbiting around the nearest star proxima Centaury.

2017 Nobelprize to Rainer Weiss, Barry C. Barish and Kip S. Thorne for the detection of gravitational waves.

2019 The Chinese spacecraft Change 4 performed a soft landing on the backside of the moon.

2020 The Mars space craft Mars 2020 landed on Mars and released the *Rover Perseverance*, which was equipped with many tools to explore the surface and a thin layer below.

2022 Start of the new James Webb Space-Telescope with greatly improved resolution and sensitivity.

Index

A
Aberration and parallax
  superposition of, 39
Age determination, 245
α-Centauri, 124
Angular momentum, 242, 243
  and mechanism, 243
Äquatorsystem, 8
Astrometric satellites
  Hipparcos and Gaia, 42, 122

B
Binaries and types, 128

C
C-Brennen, 143
Celestial body, 69
Cepheid-method and calibration, 44
Chlorophyl, 256
Chondrit meteorites, 248
Computer simulation, 244

D
Decay time, 245
Decay times, 248
Discrete line- spectrum, 128
Distance measurement and laser pulses, 40

E
Earth crust, 251
Earth moon, origin, 69
Electromagnetic pulses, 59
Elliptical orbits, 126
Extrasolar planets, 51

G
GAIA, 122
  Cepheid-method calibration, 44
Galactic center, 226
Galaxy, 250
Gaseous substances, 255
Gravitational energy, 252
Greater Bear, 122

H
Harvard classification, 128
HIPARCOS, 122

I
Ice-time, 254
Inelastic collision, 244
Inner planets, precursors of, 244

L
Light curve, 127

M
Mass-dominated universe, 184
Melting of glaciers, 254
Melting temperature, 251, 252
Molecular structures, 257
Molten metal, 251
Moon's distance, trigonometric measurements of, 40
Multicellular, 257

N
Ne-Brennen, 143
Neutron capture, 240
Non-radioactive stable isotopes, 245
Non-radiogenic lead isotope, 246
Nuclear fusion, 239

O
O-Brennen, 143
Occultation time, 124
Optical binaries, 127
Orion, 122
Ozone layer, 256

P
Parent isotopes, 246
photolysed, 256
Photometric binaries, 127
Photosphere, 128
Photosynthesis, 256
Planetesimals, 244
P-P-Kette, 97
Precession and nutation, 39
Protostars, 241
Proxima Centauris, 122

R
Radioactive dating, 250
Radio-active energy sources, 252

© Springer Nature Switzerland AG 2024
W. Demtröder, *Astrophysics,* Undergraduate Lecture Notes in Physics,
https://doi.org/10.1007/978-3-031-22135-4

Radiogenic lead isotopes, 246
Radio-isotope dating, 241
Reverse reaction, 255
Rubidium-Strontium method, 246

S
Si-Brennen, 143
Silicon mantle, 250
Solar nebula, 255
Solar system, 239, 241, 248
  origin and development, 241
Solid bodies, 249
Stars and spectral types, 128
Star spectrum, 244
Sunlight in Venus, 64
Supernova explosion, 240, 241, 244
Surface temperature, 249

T
Tetravalent atoms, 249
Trigonometric measurements, 40
Trigonometric techniques, 124
Types of Binarie, 128

U
Unstable radioactive isotopes, 245
UV radiation, 256

V
Volcanic actions, 254

W
Water vapor, 254, 256
Wavelength, 256